Computational Creativity in Media Production:

At the Crossroad of Progress and Peril

Drew R. Keller
Candidate 102
UiB Digital Kultur

14 May, 2023

Abstract

This study focuses on an approach to generate suggested video stories from raw source footage by using a multilayered hierarchical classification structure and narrative generation through the application of common patterns found within specific story genres. It frames the potential for synthetically facilitating these complex editorial decisions by analyzing current theories that define creativity and artistic quality and explores the demand for this type of engine within the creative community. The study defines the possible impact of Artificial Intelligence (AI) on artists and storytellers by examining the parallels between the effect of previous technological advancements and AI's bearing on contemporary creatives who are searching for effective modes of implementation. The potential of machine learning and artificial intelligence is rapidly advancing; therefore, this thesis traces the technology's development as a means to project future applications, specifically within media production. The Storytelling Model proposed in this report combines a multitude of algorithms that are either currently in the marketplace or exist as proof-of-concept studies. What is unique about the proposed model is the sequential and layered application of these algorithms for synthesizing a suggested video story. The recommended approach aims to provide users with the flexibility to easily modify the generated story through a prompt-feedback loop. The study's theoretical model is presented here in the hopes that it can be used as a starting point for future development.

Keywords

machine learning; neural network; artificial intelligence; emotion appraisal; story; narrative; summarization; highlights; events detection; temporal reasoning; storytelling systems; narrative modeling; story generation; computational narrative.

Audio Book

Scan to access Media

This thesis is available as an audio book. The text is read by a synthetic voice generator from ElevenLabs as an example of using Generative Artificial Intelligence for content creation.

You can also access the media from this URL: http://bit.ly/42BKDSo

Table of contents

Acknowledgements

This thesis is an outgrowth of the marriage of storytelling and technology. I have spent my life as a storyteller, spinning narratives in venues as large as a movie theater and as small as the dinner table. Now, however, my focus is on the rapidly evolving intersection of Artificial Intelligence and video storytelling, aiming to explore both creative opportunities and the social cost.

I used to scoff when I heard the phrase, “It takes a village.” It always struck me as an overused trope, a lazy description of collaboration. I now realize the phrase is trite because it is true. Crafting this thesis did take a village. My research builds on a foundation laid by hundreds of academics who have explored AI development, rhetoric, cultural analysis, and the art of video storytelling. The models and theories produced by these researchers, thinkers, and mullers provide a scaffolding of understanding, guiding me as I endeavor to disentangle the complexities of innovation. I am also grateful to those who have encouraged my academic pursuits; their contributions form indispensable pieces of this project.

Throughout my two-year tenure here in Bergen I have been fortunate to work full-time, remotely, at a global tech firm based in the United States. The support of my coworkers, supervisors, and corporate management made my academic labor possible. Before embarking on this journey my wife and I were living in Seattle, and we were measuring the benefits of moving to Norway for school. As I struggled to assess the pros and cons, I had a passing conversation with a friend, Aleš Holeček. I have immeasurable respect for his technical acumen, and his clear vision about the potential of AI. He listened to my proposal and said simply, “Do it. Just do it!” The courage he provided during that conversation has forever changed my life. I also want to acknowledge Erren Gottlieb’s contributions. As my manager at work, she listened to my ideas, ran interference with stakeholders, and during periods of academic overload she refined my responsibilities. She carved out enough room so I could meet the expectations of both school and work. Her husband, James McKenna, agreed to review my initial efforts and he eviscerated Chapter One. It needed it. He helped me find the thesis's focus and voice, ultimately shaping everything that followed. Another coworker, Doug Thomas, invested his time in reading chapter drafts, offering suggestions, and helping me ensure that the individual parts made a whole.

My academic supervisor here at the University of Bergen, Scott Rettberg, did everything right. He spurred me into action when my productivity slowed, offered insightful observations, line edited my writing, and has graciously indicated that my work didn't entirely suck. One of the most important parts of an academic journey is the collaboration and camaraderie one shares with their cohort. Each of us are exploring essential academic questions, and despite starting our studies during the pandemic, we worked to foster a strong community within the Digital Culture graduate program. Their shared laughter, their support when I struggled, and their smiles each day in our office made this journey possible. I will especially miss seeing Hanna Hellesø Lauvli (who was the one soul here in Norway who helped me navigate an opaque academic system and brought joy to the room day-in and day-out), Kira Guehring (who sat next to me, listened to me mumble, and put up with my clackety mechanical keyboard), and Florence Walker (who consistently amazed me with her insights). We were four academics from four different countries, yet we shared our love of learning.

While each and every one of these folks have generously offered their friendship, support and intellect, my deepest gratitude goes to my wife, Darcy. She courageously embraced the challenge of starting a new life in a new country, tirelessly managing the minutiae of daily life while I balanced work and studies. Most significantly in this context, she patiently proofread every word I wrote, often tackling the daunting task of deciphering rough drafts. I feel incredibly fortunate to have married my best friend, whose unwavering support made this project possible.

I want to add an acknowledgement to an unexpected resource in an academic thesis — OpenAI's Generative Pre-trained Transformer. However not just GPT, but the many AI tools integrated into most creative and productivity applications. Artificial Intelligence is now so tightly woven into the back end of these apps it is impossible to separate it from creative tools. The graphics in this thesis were created in Photoshop, which has embedded their Sensei model into the application. I have written this thesis using Microsoft's Word, which has two AI models seamlessly integrated — Editor and CoPilot. While we may not consider it as we write, when Word checks our grammar or spelling, or when Editor flags a confusing sentence, AI is the engine driving the suggestions. Additionally, much of my search for academic sources has been powered by Google algorithms. They percolate in the background, surfacing

recommendations based on my previous queries, and those of the millions of others who have searched before.

OpenAI's Chat GPT and GPT 4 have been invaluable resources for ensuring my grammar, spelling, and tone adhere to academic expectations. I have used a very specific prompt because I want it to correct errors in my writing, not fabricate it. I share my methodology here as a way to illustrate how I believe AI can expedite creativity through the elimination of the mundane.

When I have written a passage that I am unsure about, perhaps a paragraph or a phrase, I instruct the algorithm to review my work by asking, "Please act as an academic spelling and grammar corrector, applying Strunk and White's Elements of Style. List what you have changed at the end." The resultant text generated by the model is not a fabricated essay, but instead, GPT presents my work — corrected. Punctuation, spelling, tense agreement, and run-on sentences are adjusted, with specific notes at the end identifying what was changed. GPT provides me with the knowledge to make changes, and the option to accept or reject the suggestions.

The academic community is grappling with how to respond to the increasing use of AI in academic work. I recognize that this is an important conversation, and while I do not have a definitive answer to the challenge of identifying synthetic analysis passed off as original work, I believe in transparency. The AI tools I have identified above have improved the results of my creative work, and I am comfortable sharing this information because I believe I have used them thoughtfully. They have facilitated moments of clarity when my words were muddled, and I am proud to state that the ideas presented in this thesis, and the words I have chosen, are entirely my own.

0 Introduction

Video storytelling is deceivingly difficult. The process of gathering the required files, sorting through available footage, transcribing interviews, identifying important moments, writing a narrative, (with tension and release), and finally combining the parts to create a whole, is granular and time-consuming. Some video stories are clearly seen before production begins, but those projects are rare. Most often the final story will surface through editorial trial and error. It is an iterative process. The storyteller experiments with alternate combinations of story elements in search of the most effective narrative. For those who enter this alchemic process for the first time, a video story's labyrinthian development can be confusing and overwhelming.

What if a storyteller could upload their footage to the cloud, identify the look and feel of their project through a written prompt, specify the genre, provide an intended duration, and target a viewer's emotional destination? What if, moments later, a suggested story was produced from their footage? And if that story were not to their liking, what if they could refine the prompts and see an alternate version? The generated story could be the final product, but more likely, the output would be a starting point for the user to craft and refine the story in editing. This would be a process of rapid prototyping, reducing the cost of production by days. The scenario may seem far-fetched; however, much of the foundational work for assembling a computational storytelling model exists today as individual Artificial

Intelligence (AI) components. The fact that this is even a possibility signals AI is coming to media production much faster than many expected.

AI is no longer hypothetical. It inundates our daily lives. Personal assistants like Alexa, algorithm-based search on Google, and Tesla's autonomous navigation are now commonplace. This infiltration into society is unprecedented, and its reach and influence extend far beyond the scope of computing's impact during the PC and Internet eras (Eliaçik 2022). This ever-growing technology is largely unchecked by either imagination or policy, launching AI into a topic of heated discussion with both academics and the public. Much of the rhetoric is framed by either fantastical visions of utopia or dystopian worlds of singularity. While the future will almost certainly land somewhere in between these two extremes, what is common is a belief AI is bringing forth a new Industrial Revolution — one that has the potential to displace workers worldwide. What differentiates this revolution from the dawn of the machine age in the 1800s is the workers most likely to be displaced by AI are white-collar knowledge workers. Like farmhands, wheelwrights, and lace makers 150 years ago, they are facing job loss, reduced earning potential, and a realization that they must either relocate or find a new career.

One catalyst for implementing AI into media production is our voracious appetite for video content. In February 2020, YouTube, Alphabet's video platform, reported that approximately 500 hours of video were uploaded to the platform every minute. That astronomical number is comparable to over 1,200 days of video published every hour (Ceci 2023). It is but one illustration of the notion that we, as a species, are addicted to story. Even as we sleep, our minds stay active, telling us stories (Gottschall 2012). We teach through stories, we learn through stories, and we share life's pageant through the stories we tell. When we experience a story, we allow ourselves to be invaded by the teller, yielding our minds to the seduction of the tale. A story transports us through what British romantic author and literary critic Samuel T. Coleridge famously described as our "willing suspension of disbelief" (Böcking 2008). Storytelling is part of the human experience, and everyone tells stories in one form or another (Gottschall 2012).

Forty years ago, I became a professional storyteller. My video production job titles have evolved over time — writer, producer, director, editor. This professional experience brings a realization that the advent of Artificial Intelligence will usher in tectonic changes, shaking the foundations of my craft.

How this technology gets implemented is still in a state of flux. Will the emphasis be to build AI functionality that contributes to enriching the human creative space? (Arriagada and Arriagada-Bruneau 2022) Or will AI render many of the multitude of crafts within media production obsolete? Graphic designers, writers, editors, and composers already face potential obsolescence with the development of applications like Stable Diffusion, Jasper, Synthesia, and Soundraw. Researchers have expressed a belief that AI will eliminate many of these human jobs (Hong and Curran 2019). Zach Sutton writes, “The ability to create concept art, storyboard an idea, or bring concepts to life is now easier than ever – and those jobs will suffer dramatically from this, just as the painters hired to paint portraits did in the 1800s” (Sutton 2022, 13).

As individuals, we leverage inspirational references when we create, drawing from what we have seen, experienced, and know. The act of creation is combining these sources into something new. A catalyst for the trepidation expressed by creatives is AI's similar capacity to combine referents and generate output. Moreover, the algorithms perform this task at scale, bringing the cost down to near-zero (Ramesh 2022). To compound the threat, viewers often find the generated output engaging. Researcher Alan Thompson believes AI generated illustration is now virtually indistinguishable from that of a human (Thompson 2022).

This rush to AI will confront us with decisions about who, or what, creates our stories. The paradigms of creation are shifting, altering not just the techniques of storytelling, but redefining the message itself. This tension is not new. In an article that explores how reproduction alters art, Walter Benjamin quotes Paul Valéry's observations on technology's impact on early twentieth-century artists. In *La Conquète de l’ubiquité* (1928) Valéry observed, “We must expect great innovations to transform the entire technique of the arts, thereby affecting artistic invention itself and perhaps even bringing about an amazing change in our very notion of art” (Benjamin 2006, 18). Elaine Scarry believes this change can come at both a cultural and personal cost, noting, "We make material artifacts in order to interiorize them: we make things so that they will in turn remake us, revising the interior embodied consciousness" (Freedgood 2003, 641). Certainly, the innovations brought by AI will be transformative, particularly in video production. However, the risk of innovation comes with losing the multitude of personal video stories crafted from a human point of view. It will mean abandoning the artifact of creation— a human process of examination and exploration.

Computer and Cognitive Scientist Margaret Boden writes that creativity is often seen as the pinnacle of intelligence, humanity’s crowning glory. And yet, it is not fully understood (Boden 2004). She describes creativity as the capacity to generate ideas or outputs that are unique and worthy, which are "new, surprising and valuable" (Boden 2011, 29). Creativity is described as a moment of inspiration, a “mysterious muse” that arrives like a bolt of lightning, a burst of insight and imagination (Arriagada and Arriagada-Bruneau 2022, 81). As Keith Sawyer describes creativity as “part of what makes us human” (Sawyer 2012, 3).

This thesis, at its core, is an examination of whether machines are capable of creating things that are comparable to what a human creates. When they generate content that appears to present emotion, imagination and insight, fabricating content that resonates with a viewer, are they then the creator? Is computational output an example of creativity?

Demonstrations of creativity are often inscrutable, making it difficult to quantify the act. One useful definition of creativity is the capacity to develop new, novel, and valuable ideas in a surprising or unfamiliar way (Kurt 2018). Creative ideas appear in a multitude of forms, from concepts to theories, from paintings to architecture, from music to stories. We are all capable of creativity, and today it is perceived as a part of everyday life and a key feature of human intelligence. However, Robert Paul Weiner observes, "The word 'creativity' did not exist before 1870 and was not widely used until about 1950" (Weiner 2000, 15). He believes that contemporary concepts of creativity, "and the positive value we attach to it, might in fact be seen as hallmarks of our modern, secular, democratic, capitalist society. What we mean by creativity seems to be shifting" (ibid. 2000, 15).

While human creativity has a large corpus of study, there is less agreement about what constitutes computational creativity. One approach is to leverage the same evaluative scaffolding that is used in the appraisal of human-created visual art – schema theory. The proponents of AI's ability to generate artistic creativity argue that AI models "appear" to be creating art, therefore it is creative (Arriagada and Arriagada-Bruneau 2022). In the spirit of René Descartes "cogito, ergo sum" (I think, therefore I am), the belief is if it looks like art, then it is creative. Criticism surrounding AI's ability to foster true creativity often stems from the notion algorithms are void of important human traits. They cannot be actual creative agents because of the lack of personhood, cultural integration, and authorship (ibid. 2022).

The human's role is one component when measuring the effectiveness of an artistic artifact's ability to convey meaning. After all, humans evaluate the work; be it a painting, a poem, or video story. We measure the perceived value and relevance within our personal scales of taste, aesthetics, and experience. For a creative work to carry relevance and value, the effort should be relatable for a human audience (Kurt 2018). Caterina Moruzzi believes it is important to consider an audience's intuition and response as the determination for measuring AI's creativity (Moruzzi 2020). Following this line of thinking, creativity of any form, be it human or machine, will ultimately be judged by humans (Arriagada and Arriagada-Bruneau 2022).

For many people, the explanatory process of creation is critical in appreciating creative work. They want to understand the "theory of mind" of the creators (Das and Varshney 2022). They want to grasp the artist's motivations and intent. Unfortunately, applying this metric of appreciation to AI-generated media poses an inherent challenge. The process of creation within these models is so opaque that even the developers themselves are not entirely certain how they work (Ouyang, et al. 2022). Many discount the results of AI creativity out of hand because of this lack of procedural provenance.

As AI is being integrated into all facets of creativity, the question arises: if storytelling is an expression of the human experience, how will AI-generated content change our perception of ourselves? It seems humans are in the process of redefining what we consider to be demonstrations of our 'humanness' If stories help us to position ourselves in society, will there be consequences when stories are not an expression of personal examination and human exploration, but instead an aggregation of thousands of stories that have come before? What is the potential and peril we face with machine-made media? These are the questions that have shaped the focus of this thesis: <u>How might Artificial Intelligence impact video storytelling, and what is the potential for incorporating AI automation into media creation?</u>

Method and Approach:

Because of Artificial Intelligence's complexity, I have chosen to work sequentially through five issues that frame both development and deployment. To address the research questions, this study will contain five chapters and a conclusion:

- What is the impact of automation on creativity?
- Do previous eras of technological change inform us about AI's potential impact on creative workers?
- How does the historical development of AI frame contemporary applications?
- What models are likely to be leveraged for video storytelling?
- A proposed workflow for creating video narratives with Artificial Intelligence.

For a comprehensive understanding of this technology, Chapter One introduces the concept of Artificial Intelligence by defining human and machine intelligence, discussing creativity and its measurement, and exploring the rationale for creating specific AI models. It also examines the astronomical investment required and questions the value assigned to human versus machine creativity. Additionally, it considers the potential consequences of replacing our culture's storytellers with algorithms. Chapter One aims to provide a technical foundation for comprehending Artificial Intelligence, as well as to identify key societal questions, forming the framework for further explorations in this study.

Chapter Two explores the parallels between the current expansion of computational creativity and the effects previously experienced by workers during four industrial revolutions. It examines the history of lace manufacturing in the 19th century as a means of gaining insights into the societal consequences of technology undermining a creative workforce. AI's present implementation seems to be targeting the jobs held by college-educated workers. Consequently, the chapter identifies some of the potential challenges these workers may encounter when they face lost wages and difficulties in career advancement due to an oversupply of unemployed and skilled professionals.

Chapter Three traces the development of Artificial Intelligence, beginning with the mapping of the human neuron, the work of Alan Turing, and continuing through the efforts to develop a working demonstration of computational intelligence. It outlines the differences between symbolic and neural models, and explores the differentiation between Narrow, General and Super AI. Additionally, it includes how creative workers are experiencing shifting occupational demands, their anxiety about AI’s potential to marginalize their creative efforts, and how some are finding paths that differentiate their work from machine-generated media.

Chapter Four is a technical examination of the generative models most likely to be implemented for computational video creation. It explores the innerworkings of Generative Adversarial Networks, Creative Adversarial Networks, Diffusion Models, Contrastive Language–Image Pre-training (CLIP), and how these models are being applied to image, sound, text and video generation.

Applying the knowledge of generative networks gained in Chapter Four, Chapter Five presents a proposed Storytelling Model. The aim of the algorithm is to generate suggested video stories from raw source footage by employing a multilayered hierarchical classification structure and narrative generation based on common patterns found in specific story genres. The model facilitates these complex editorial decisions by initially training on specific genres of video stories and conducting image sentiment analysis, and then applying this knowledge to source footage guided by a user's prompt.

The final chapter is a summation of both practical and potential outcomes from incorporating algorithms of creativity into video production.

This thesis will explore contemporary developments in Artificial Intelligence within a framework that leverages cognitive science, computer science, history, phenomenology, and media studies. It aims to provide an understanding of generative AI media, the capabilities and limitations of these models, and the potential harm and benefits to both individuals and society.

My research was motivated by a personal desire to better understand the rapid changes we are experiencing with the advent of AI. Over the past eighty years, the technology has been on a gradual upward trajectory of development and adoption. Useful technologies generally follow a sigmoid curve, as noted by Hipkins and Cowie (2016). Adoption begins slowly with the initiation of invention and the introduction of a "new thing". This phase is where pain points, obstacles for adoption, and rapid iteration occur. As people discover the technology's potential, there is an explosion of growth, with adoption rates ramping up the curve. This critical mass of non-technical users drives more innovation, capital, investment, and competition, rapidly accelerating the implementation of new uses. It becomes a race for new, better, and faster uses for the technology. The technology disrupts industries, alters how we do our jobs, and changes how we live. A prime example of this disruption would be the internet's impact on the music industry in the 1990s. Eventually, as a technology matures, the rates of development and adoption slow, and the technology becomes a resource woven into our day-to-day lives. From electricity to telephones to the internet, all have followed a sigmoid curve.

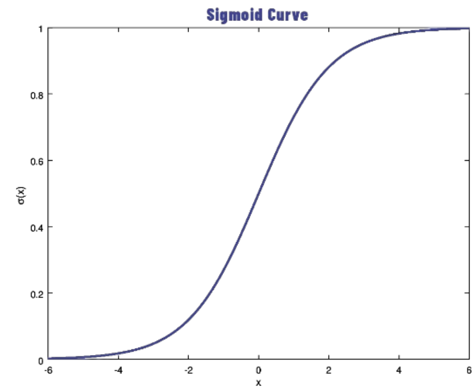


Figure 1 Sigmoid Curve

The storm surrounding artificial intelligence in popular culture, specifically in the first five months of 2023, may well represent the first rumble of thunder signaling that our lives are about to change. If the launch of AI is more than media hype, then it appears that the shift we are about to witness will be rapid. It is likely this thesis will quickly become dated, as it documents a particular point in time. The daily announcements of new developments, research, and complications will quickly date the technology described in this thesis. However, the context provided on the development, implementation and impact of AI should be relevant for much longer.

As a species, we continue to create. The process of making things is fundamental to what it means to be human. Mark Hatch, the owner of the TechShop chain of Maker studios wrote, “We must make, create, and express ourselves to feel whole” (Hatch 2013, 11). However, will AI-generated stories provide a sense of personal exploration and authorship for the creator? My existential dread revolves around the personal adaptations required of me for continued occupational relevance. These tools will generate different types of stories, but the question I pose is, will this new paradigm create stories that resonate with the viewer, or will they be disposable artifacts? Valéry’s observation that technology is often the catalyst for an “amazing change in our very notion of art” could prove prophetic.

Chapter 1

What is the impact of automation on creativity?

1.1 "More profound than fire, electricity, or the internet" - Sundar Pichai, Google

We are surrounded by technology. Our inventions are intricately woven into the fabric of our lives, from what we wear[1], what we carry in our pockets[2], how we move from place-to-place[3], even what we eat[4]. These are ever-shifting forces whose constant iterative changes force us to adapt or disengage. Technological change has been driven by our human desires (Roose 2022). The stone tablet, the wheel, the steam engine, social media — these things didn't appear from the ether, fully manifest and integrated into society. They are the product of human thought and design.

With four decades of experience in video storytelling, the constant drumbeat of technological change has forced me to cultivate a mindset that emphasizes reinvention. In that time my workflow has undergone radical transformations; from analogue tape to digital files, from a Steenbeck film table to computer-based video editing, and from art cards to motion graphics systems. Mastering the latest camera, recorder or design platform is a function of occupational sustainability. My personal experience has prompted me to cast an eye over the technological horizon and ask, "How will the implementation

[1] Spun bonding Technology and Fabric Properties: A Review (Midha and Arjun 2017)
[2] Use That Everyday A.I. in Your Pocket (Biersdorfer 2022)
[3] On a Formal Model of Safe and Scalable Self-driving Cars (Shalev-Shwartz, Shammah and Shashua 2017)
[4] Genetically modified foods: safety, risks and public concerns—a review (Bawa and Anilakumar 2013)

of AI into video production impact visual storytelling? What will this new wave of storytelling look like? Will this AI revolution displace creatives like me?”

Artificial Intelligence has, and will continue to assume, a greater role in human activity. Since its emergence, AI has sparked debate about whether it will be the key that unlocks human society's success or lead us to ruin (Eliaçik 2022). Proponents believe it has the potential to unlock abundance from scarcity, including human creativity (Abram 2022), thus revolutionizing the act of creation (Tufekci 2022). Developers tout new capabilities brought about by this technological shift. Sundar Pichai, the CEO of Google, asserts that AI “is more profound than fire, electricity, or the internet” (Steiner 2021).

One rationale for increasing the role of AI in creative paths is when you replace a human with a machine, you can, in theory, free that human up to do more valuable things (Roose 2022). Furthermore, we are witnessing the application of Artificial Intelligence not just as a creative tool, but as the actual artistic creator (Cetinic and She 2022). The emergence of machines displaying creative agency presents an important question: Can a machine think and create like a human? For many, the baseline measurement and ultimate goal of Artificial Intelligence is to mimic and eventually surpass all forms of human intelligence.

To frame this exploration on the impact of AI on video storytelling, it makes sense to first define human intelligence.

Human general intelligence is described as the mental ability for reasoning, problem solving, and learning (Colom, et al. 2010). Over our lifespans we can reason through countless issues, acquiring simple and complex behavioral repertoires. While each of us has differing capabilities for reasoning and learning, human intelligence affords us the ability to cope with challenging situations (ibid. 2010).

The definition of Artificial Intelligence leverages much of the same descriptive framework — reasoning, problem solving, and learning. Deniz Kurt describes AI as, “the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision making, and translation between languages” (Kurt 2018, 9). This definition of AI identifies the computer model’s capacity to ‘think’ as it learns, solves problems and reasons out a solution.

Connecting the word ‘artificial’ to the word ‘intelligence’ delineates that the results of this intelligence are generated by a computational system, or as cognitive and computer scientist John McCarthy defines AI, it is the “science and engineering of making intelligent machines, especially intelligent computer programs” (McCarthy 2007, 2). He amplifies this point with the observation that for both people and computers, demonstrated intelligence is the “computational part of the ability to achieve goals” (ibid. 2007, 2).

While AI and human intelligence share a common descriptive framework that forms a baseline for comparing both the successes and failures of AI implementation, there are limitations. McCarthy contends that due to the inherent complexity of human intelligence, computers should be deemed only “somewhat intelligent” because they use linear logic to complete specific tasks (ibid. 2007, 3). As we will explore in greater detail in Chapter Three, specifically through the writing of Alan Turing, the demonstration of computer intelligence should be more than just answering a binary yes/no question. The consideration of AI as only “somewhat intelligent” is important because of the limited scope of their capacities. These models are programmed to accomplish specific tasks, narrowing their abilities and limiting them to the replication of specific elements of human intelligence, often leaving them incapable of performing others. Furthermore, McCarthy argues that we cannot clearly define artificial intelligence until we have a clearer understanding of human intelligence.

The human brain is a marvelously complex and enigmatic system. To understand how human intelligence differs from a computer, it is helpful to contrast the differences. Research shows that our brains are highly complex organs that perform a vast array of functions through neurotransmitters. This includes sensory processing, motor control, and cognitive processes (Colom, et al. 2010). While the brain's analog domain results in a system that is inherently sluggish at processing calculations, Ray Kurzweil points out that most neurons work in parallel, executing up to one hundred trillion computations simultaneously (Kurzweil 2014). This number of computations may seem unfathomable but pales in comparison to the current 500 qubits of processing power of contemporary AI quantum computing, with expectations that by 2025 quantum processors will scale to over 4,000 (Lardinois 2022) (Colbrook, Antun and Hansen 2022).

While the human brain has comparatively limited analog power compared to a quantum computer, one defining part of our intelligence is our capacity to recognize patterns through random analysis (Kurzweil 2014). Our ability to perform these random interactions has been described by Kurzweil as a "chaotic dance," where our neural system performs massive parallel computations. In contrast, existing AI systems lack the ability to leverage the arbitrariness and random interactions of the human system, working to solve a problem via linear brute force. This limitation to perform random pattern matching, in essence an inability to mimic the human brain, is a significant limitation of artificial intelligence (Kurt 2018).

Even with these limitations, AI has become an integral part of everyday life. We experience implementations of Artificial Narrow Intelligence (ANI) leveraging Big Data and complex algorithms in our voice assistants, our smartphones, and the commodification of attention in our social media timelines (Pedersen, Albris and Seaver 2021). Previous digital creative technologies like Computer Aided Design (CAD), Adobe's photo, video and illustration applications, and motion capture systems have recently implemented AI enhancements. A graphic designer who previously created illustrations in Photoshop with hand drawings, geometric shapes and images can now generate unique artwork directly inside the application via Adobe's Firefly neural network (Adobe 2023). These implementations of AI in the creative process are seen as stimuli for artists (Arriagada and Arriagada-Bruneau 2022). Artists and researchers are touting AI as a path for not just supporting artists, but "creating art as re-embodied via computational abstraction processes, and actually make new forms of art" (Cheng 2022, 5). They see AI as a path for unlocking the human psychological process of creation.

AI is revolutionizing the artistic sphere. This integration of Artificial Intelligence as a new and unique stimulus for artistic creation will have significant implications for both creative output and the work of creators (Arriagada and Arriagada-Bruneau 2022). It is forcing us to ask ourselves, what is creativity? Is it an inherently human trait, or is it something that can be bottled and replicated by a machine? What will be the role of creators and artists in this era of computational evolution, and how can they compete with powerful technologies that operate on a vast scale with enormous computing capacity (Roose 2022)? To begin to address these questions, it is worth examining how technology influences creativity to gain a better understanding of the potential impact on both creators and their creations.

1.2 AI in everyday life

Invention and creativity are both iterative processes – the results are built on the foundation of previous work. Through history each of our technological leaps are really the culmination of hundreds, if not thousands, of small preceding technological steps. We see examples of this iterative development in the tools and devices we leverage daily. For example, when camera autofocus was first introduced in 1985 it was an unreliable and often sketchy assistive technology. As a photographer, I found the camera would lose focus on the intended subject and start focusing on random elements in the frame. Trees in the background could be clearly seen while the people in the foreground were wildly out of focus. Consequently, most professional photographers realized it was best to disable autofocus to ensure their images were sharp (Edwards 2022). After 35 years of refinement and reinvention, including infrared sensors, facial recognition, and phase detection, autofocus has iterated into a reliable tool trusted by both amateurs and professionals alike.

We are observing a similar iterative path in the development and implementation of AI. This process of technological evolution can be illustrated by OpenAI Labs' large language model. When it was first unveiled in 2018, their natural language application, Generative Pre-trained Transformer (GPT), was regarded as being superior to existing language algorithms. This was particularly true for tasks such as common-sense reasoning and reading comprehension (Hajj 2023). GPT was a natural language processor that could interact via text prompts, engaging users in what appeared to be a natural conversation between the AI and the user. GPT seemed to understand sentences and effectively reason through ideas. Eight months later, OpenAI released a larger version called GPT-2. This version was trained on over ten times the amount of data, imbued with the ability to generate more natural-sounding text (ibid. 2023). In June 2020, OpenAI released GPT-3, followed in November 2022 by a more user-friendly iteration, ChatGPT. ChatGPT is a prime example of generative artificial intelligence, a computational system capable of producing unique text, visuals, sound, and other media in response to short prompts (Metz and Weise 2023). Zeynep Tufekci described ChatGPT as much more than "another entry in the artificial intelligence hype cycle" (Tufekci 2022, 1). She writes that ChatGPT is capable of generating essays in response to open-ended questions, all delivered with the voice of a knowledgeable high school essay. In a New York Times article, Frank Bruni described ChatGPT as "a surprisingly competent writer and sometimes even a clever one, to the point where early users regard it as 'some mix of software and sorcery'" (Bruni 2022, 1). Each subsequent GPT version was trained on significantly larger data sets, using considerably more computing power and leveraging more refined algorithms (Hajj 2023). ChatGPT is the result of developing a tool that gradually became more efficient, more interactive, and more powerful.

ChatGPT was built on a data set of text curated from the internet before 2021 (Thompson 2022). The process of training a large language model involves teaching it to recognize relationships between billions of words and generate a representation, or token, for each word, syllable, or phrase. These relationships are analyzed to identify usage patterns, common combinations, and language structures. When ChatGPT generates bespoke sentences and paragraphs as it fabricates an essay, the algorithm does not access the original information. Instead, it applies what it has learned from its training by predicting the most likely next word in a sentence. This approach models the way the human brain performs the same task, applying its acquired language patterns and subject knowledge. One significant benefit of GPT's computational approach of applying predictive patterns rather than retrieving specific

	Wikipedia	Books	Journals	Reddit links	CC	Other	Total
GPT-1		*4.6*					4.6
GPT-2				**40**			40
GPT-3	*11.4*	*21*	*101*	*50*	**570**		753

Figure 2 **Summary of Major Dataset Sizes**. Shown in GB. Disclosed in **bold**. Determined in *italics*. Raw training dataset sizes only. *(Thompson 2022) Alan D. Thompson. 2022. What's in my AI? A Comprehensive Analysis of Datasets used to Train GPT-1, GPT-2, GPT-3, GPT-NeoX-20B, Megatron-11B, MT-NLG, and Gopher*

passages of text is that training a large language model is both time- and energy-consuming, often taking months to complete. If applications like ChatGPT were required to parse billions of lines of text in their datasets as they generated a response, they would, for all intents and purposes, be retraining themselves. The process would be a computational bottleneck as the source data is huge (*see figure 1*). By not parsing the original data and disconnecting the application from the source, programmers save compute time and reduce cost. While this approach is more efficient, the downside is that the trained model is neither contemporary nor dynamic. ChatGPT users often encounter dated responses to topical prompts. Many generated responses include content that does not incorporate contemporary information or current events because the LLM is simply unaware that the information exists. Some instances of the generated output present facts that are sporadically inaccurate and occasionally wildly wrong (Roose 2022). The prompt page on the Open AI website even warns users that ChatGPT, "may occasionally generate incorrect information" (Open AI 2022).

Recent studies have shown that these generative Natural Language Processing (NLP) models are prone to "hallucinate" incorrect content (Zhou, et al. 2021). In more colloquial terms, they can make stuff up. An NLP hallucination is plausible sounding nonsense where the output generated deviates from facts or factual logic. The scale of this nonsense can range from sentence contradictions (where the information in a sentence does not agree with the previous one), prompt contradictions (where the response diverts from the prompt request), factual contradictions (where the model fabricates incorrect information), and nonsensical hallucinations (information that is completely irrelevant) (Keen 2023). For researchers, there is no obvious explanation for this misaligned production. IBM scientist Martin Keen indicates that it is difficult to find a solution "because we still don't entirely understand how these models derive their output" (ibid. 2023, 3:49).

A likely contributor to the phenomena of hallucinations is that these sophisticated NLP models are trained to rephrase and summarize long and often labyrinthine tracts of source text. The hallucination problem arises because most current models do not have an architecture that encapsulates specific information in context (Anderson 2022). Therefore, individual facts are nothing more than tokenized blocks, and an NLP reassembles those non-contextual tokens into what it has been trained to believe is the correct output.

To further compound the problem, if the data sourced from the internet and used for training includes text files with faulty analysis or bad information, ChatGPT has no mechanism to vet the accuracy of the source. This is particularly true when a model is trained on a large corpus of text that contains noise, errors, biases, or inconsistencies. For example, Wikipedia and Reddit are common sources in these datasets, and they are not known for rigorous fact checking and vetted research (ibid. 2023). ChatGPT assumes the source information is correct, incorporating the misinformation in the generated report (Chen 2022). Algorithms are only able to identify things that are quantifiable and are, therefore, incapable of making subjective judgments, judgments like quality or accuracy (Broussard 2018). The resulting output presents a perceived authority through a sort of contextual osmosis (Anderson 2022). For users, unless they are well-versed in the essay's topic, they could be subject to what Tufekci describes as a "high-quality intellectual snow job. You would face, as Plato predicted, 'the show of wisdom without the reality'" (Tufekci 2022, 1). Or, as video educator Jordy Vandeput described his experience, “AI isn’t perfect yet” (Vandeput 2022, 00:26).

The appeal to users of applications like ChatGPT is they afford an expeditious shortcut to the sometimes-difficult task of writing. One concern raised by educators is when students overly rely on AI writing tools, the art of crafting an essay could possibly go the way of cursive writing or ballroom dancing — skills practiced by only a few. Tufekci believes that by allowing the current generation of students to curtail the development of essay writing skills, a task that requires research, comparative analysis, synthesis, and developing a clear point of view, the loss could impede their intellectual progress (Tufekci 2022).

An additional complication with generated text flooding our information streams is the probability that we will need to develop personal content filters, refining our ability to discern truth from “plausible but incorrect text”. The outcome of users implicitly trusting AI content may be analogous to how David Beer described the deleterious impact of social media algorithms on our culture (Beer 2013). These AI generative algorithms will be shaping how people internalize what they have learned, how they embrace contemporary culture, how they perceive truth, how they create new work, and how they communicate their views to others.

The challenge may be more than simply maintaining our humanity in a world that is increasingly designed for and by machines. It may also be understanding and redefining creativity as we integrate new and unique stimulus for artistic creation. We have always defined creativity from the perspective human creativity (Arriagada and Arriagada-Bruneau 2022). So how can we evaluate the emerging role that AI is playing in the creative process?

1.3 AI in the creative space

The inner workings of these AI tools are shrouded in mystery. While developers have a clear theoretical understanding of the transformer model within an application like GPT, when algorithms demonstrate an ability to acquire new skills without specific training, present a sophisticated application of knowledge, or generate unexpected results, researchers find it difficult to quantify the behaviors. IBM's Martin Keen and OpenAI's Long Ouyang describe the innerworkings as enigmatic, with developers uncertain about how these models arrive at their results (Keen 2023) (Quyang 2023). In a research paper published by the Open AI Alignment team, a part of the cohort that developed GPT, developers admitted that their model is almost a black box and that, "We don't exactly know what it's doing" (Ouyang, et al. 2022). An additional complication is the development of many AI models is not occurring in controlled and closed lab environments. Instead, the public is performing most of the large-scale testing as companies make beta versions of the applications generally available. Updates and improvements with these tools are current and dynamic. Even Jack Clark, Open AI's head of policy, was quoted in an interview saying, "We're trying to build the road as we head across it" (Hern 2019).

Yet, to be fair, if the measure of success for an AI model is its comparison to human intelligence, then that metric is flawed as many of the mechanisms of human intelligence are also a mystery to researchers (McCarthy 2007, 3). It becomes difficult to measure the effectiveness of AI's intelligence metrics when researchers are unable to define a baseline process for human intelligence. Because this comparative metric is fuzzy, it is nearly impossible to empirically measure AI's success or failure, risk and reward, or even computational competency.

These fuzzy success metrics have not deterred financial analysts from extolling the potential of AI. Investment analyst Will Summerlin writes that he believes AI models like Open AI, DeepMind, and others are demonstrating near human-level proficiency in many knowledge worker tasks (Summerlin 2022). Sonya Huang and Pat Grady, from Sequoia Capital, describe a belief that as AI models have become progressively larger, they have surpassed human performance benchmarks (Huang and Grady 2022). Although these analysts specialize in finance rather than quantitative computing, their enthusiasm illuminates a rapidly expanding AI landgrab.

A catalyst for many of these breathless evaluations by financial analysts is the inclusion of generative AI in existing creative software ecosystems. Currently, software developers complete code using AI-powered integrated development environment (IDE) software, designers generate graphics in Figma and Photoshop, and Discord bots inject generative AI text into social communities (ibid. 2022). These implementations are deployed as application modules and plug-ins, connecting a user to vast AI ecosystems. Huang and Grady describe these integrations of AI as a "little brain" that sits on top of a large general-purpose model "big brain". The value to a user is that these implementations of generative AI live within an environment of a known software application. The capability of the software expands significantly by integrating AI's ability to process large sets of data and perform pattern analysis, with results incorporated into the user's workflow. These implementations are presented as a way to augment human creativity through speed, efficiency, and the presentation of unexpected solutions.

1.4 What is “good”?

The argument extolling the benefits of speed, efficiency, and unpredictability is an oft-heard refrain from corporate marketers, investors, and developers. However, this push to ramp up worker efficiency appears to be counterintuitive when it comes to creativity, a classic contemplative and internal process. Is speed and efficiency of value to creatives? When AI becomes a cornerstone in creativity, how do we decide if the result is comparable, or even better, than an example of human creativity? Are the results good? And what constitutes "good"?

For a thousand years painting has been considered the pinnacle of human creativity, representing humanity's purest expression (Hong and Curran 2019). However, how do we perceive the merit of art when it is created by a machine? Can it be deemed to display inherent artistic intelligence and creativity? In their paper “Artificial Intelligence, Artists, and Art: Attitudes Toward Artwork Produced by Humans vs. Artificial Intelligence” (2019), Joo-Wha Hong and Nathaniel Curran raise the question whether contemporary advances in AI-generated artwork complicate our understanding of creativity and aesthetic beauty. When machines exhibit artistic abilities, when they execute a creative process, is it truly creative? Mingyong Cheng put it in simpler terms, “Can machines create art (Cheng 2022)”?

Computer scientist Donald E. Knuth wrote, “Science is what we understand well enough to explain to a computer. Art is everything else we do” (Knuth 1996). Expressing our humanity through art is a part of every culture and clan. Keith Sawyer posits that creativity is “part of what makes us human” (Sawyer 2012, 3).

In 2018, the French artist-collective Obvious made headlines when their AI generated painting "Edmond De Belamy" sold for a record $432,000 at Christie's art auction (Whiddington 2022). Their portrait painting leveraged a Generative Adversarial Network (GAN), referencing 15,000 portraits painted between the 14th and 20th-century, to generate a unique piece of work (*figure 2*). The sale was noteworthy because of both the purchase price and the novelty of a computationally created artwork being considered a “legitimate” piece of art. However, at the time of the sale, there were questions about who should receive credit for the work: Obvious or the artists of the original source paintings? Additionally, researchers wondered if the anthropomorphic projection that the painting was created by AI contributed to the perception of artistic merit and the artwork's success (Epstein, et al. 2020).

Figure 3 "Edmond De Belamy",
Obvious AI & Art Collective, (Obvious 2017)

If artwork is defined as a representation of human intellect and creativity, is it still a demonstration of creativity when it is generated by algorithms instead of paintbrushes? One argument is that when a human uses AI to create art, it becomes a symbiotic relationship between art and technology, an illustration of the ever-evolving correlation between human intelligence and machine intelligence (Kurt 2018). This computational creativity represents a melding of humanity and technology. However, there are identifiable differences between human and machine creativity. When defining the uniqueness of the human brain, creativity is often presented as the discriminator that highlights those differences (Cheng 2022). Creative acts represent the aesthetic and emotional capacities that define us, they are a sophisticated web of cause, effect, and happenstance.

As humans we display our creativity through the mental process of combining associative elements into a new form (Das and Varshney 2022). The way we think, recall and process information are the combinatorial mechanisms we use when we create associations that connect two or more concepts. Creativity has been described as our ability to solve problems or create something useful by forming associative combinations of new and existing knowledge (Sawyer 2012). Keith Sawyer believes that for an individual, “creativity is a new mental combination that is expressed in the world” (ibid. 2012, 7). Margaret Boden differentiates creative ideas into psychological and historical acts of creativity (Boden 2004). She sees psychological creativity as an evolution of unpredictable ideas, ideas which are new to the individual. If an individual forms an idea that is new to them, even if the idea has been previously conceived by others, it is still a demonstration of creativity. She distinguishes historical creativity as the resulting artifact where an individual has formed a unique idea that has never previously been presented (ibid. 2004).

For a creative act to be recognized, be it psychological or historical, it needs to be expressed, be it spoken, performed, written, drawn or shared. Sawyer writes that “creativity is expressed in the world” (Sawyer 2012, 7) because if an idea is only internal, neither shared or expressed, it is therefore neither seen nor understood. This implies that a conceived idea needs to be presented to receive feedback (Cheng 2022).

Creativity can also be viewed through a framework of culture. Sawyer writes that ideas have a cultural point of view and are measured within a community that understand and appreciate the ideas. “Creativity is the generation of a product that is judged to be novel and to be appropriate, useful, or valuable by a suitably knowledgeable social group” (Sawyer 2012, 7). This implies that judgements of creativity can be fluid, as cultural knowledge and understanding are critical components when assessing merit. Berys Gaut writes in “The Philosophy of Creativity”, that an assessment of creativity depends on “judgements of a field of experts using the appropriate standards of the historically conditioned domain of activity” (Gaut 2010, 1038). These individual, social, and cultural influences can be the driving force that shapes the development of an idea. Therefore, when judging the creativity of an idea it is important to consider how imitation and societal tradition are imbued in the product, as they provide context.

How then, might these definitions of human creativity, and judgements of effectiveness, apply to AI-created media? Are there approaches to identifying when AI output contains features and values that indicate their creativeness? Cheng leverages Boden’s three forms of creativity to argue that AI is in fact creative: combinatory, transformational, and exploratory (Cheng 2022).

Combinatorial Creativity is both iterative and associative. It is the application of existing ideas or artifacts brought together in a new way (Das and Varshney 2022). The "newness" presented in generative media's combinations is not solely driven by the application of AI's brute computing force. Each instance of text, image, or sound is unique due to an inherent randomness integrated into the generative algorithms as a mechanism for improving the fidelity of the final output. The model accesses, synthesizes, and presents fragments of the original sources in an entirely new light. These generated artifacts are created as binary strings, something N. Katherine Hayles would classify as "digitally born" (2016). The combinatorial creativity presented in this media represents the most recent link in a long chain of combinatorial poetics. These algorithms synthesize sources that reflect humanity's individual, social, and cultural artifacts, producing a unique digital amalgamation of ideas, with the likelihood the instance is destined for the digital ecosystem.

Cheng argues that transformational creativity is represented in generative media through completely new interpretations of a conceptual domain. By transforming an abstract idea or space, entirely new ideas or concepts are generated that were previously unimagined (Cheng 2022). Imagination, in fact, is the critical component in triggering this transformational creativity.

Exploratory Creativity involves producing novel ideas within the adopted standards and styles of a particular school of thought. These instances of creativity are unexpected and new, and the exploration needs to be a consistent representation within that accepted school. Therefore, appreciation of these ideas requires a recipient who understands the cultural framework of the style (ibid. 2022). According to Cheng, at its core, machine learning is an execution of Exploratory Creativity because Artificial Intelligence provides a structure of algorithmic styles that are learned and then implemented into new concepts. By leveraging an existing style of thought, a user can direct an AI model to apply that specific style to a concept, creating a unique outcome that fits within expected standards.

Applying the definition of human creativity established earlier, whereby an idea or artifact is considered a display of originality by a knowledgeable social group if it is novel, useful, or valuable, we can distinguish between the two primary classifications of AI image models — Generative AI and Creative AI. As a form of combinatorial creativity, Generative AI algorithms are interpolative rather than extrapolative, synthesizing source data into artifacts that reflect a specific prompted style. Creative AI is a subset of generative AI, intentionally designed to output novel and unexpected output. These algorithms are referred to as Exploratory Creativity algorithms, and they model the act of human inspiration by explicitly attempting to extrapolate beyond their training set to produce unimagined ideas or artifacts (Das and Varshney 2022).

While these models generate previously unimagined images, are they relatable to humans? Do they have artistic value? Can the products of AI be accepted as equal to the ideas created by human artists? Computers do not have taste or aesthetic sensibilities; they cannot judge if the products they create are interesting and sensible. What they produce could be considered a masterpiece or a nonsensical jumble. For the result to have value to a viewer, it needs to express a human point of view that resonates.

Traditionally, the purpose of art is defined as the communication of an idea between individuals, an idea that needs to be both explainable and understandable. In other words, it should convey human intent, inspiration, or expression (Collingwood 1938). This communication of meaning is necessary for stimulating an aesthetic experience with the viewer (Das and Varshney 2022). However, how is AI-generated art received when the product does not reflect an identifiable human point of view? Images generated by AI diffusion models are referencing sources that were originally created by humans, so logically one can consider them as trained to mirror human expressions. The intended outcome of synthesizing human sources into new media is the generation of art that resonates with a viewer. If the classification of "art" depends on the subjective judgment of a human, if someone finds it appealing then logically anything can be classified as art, from Rembrandt to a velvet Elvis to AI. While an AI model does not possess a specific point of view, as it is merely parroting the style and aesthetics of the source data, when a viewer responds to the produced media it should be considered art. Mark Coeckelbergh argues that AI-generated media should be included within the defined framework of creativity and art because it is appealing and therefore meets the objective and subjective criteria of art (Coeckelbergh 2016). As this classification means most anything can be classified as "art" (qFiasco 2018), perhaps a better question is, can AI create art that is good and worthy?

Beauty may very well be in the eye of the beholder because a viewer's bias will be a significant contributing factor when they determine the creative merit of AI-generated artwork. Viewers who hold a stereotype that AI models are incapable of "humanlike" performance, even when the resulting output is objectively indistinguishable, are unlikely to find value in AI-generated artwork (McCarthy 2007). Therefore, these evaluators believe that because artwork can only be created by humans, AI cannot make art. Conversely, to find attractiveness in AI-generated work an evaluator needs to hold a belief that AI can be creative (Chamberlain, et al. 2018). Removing these biases from the evaluative process can prove difficult. Hoo-Ha Hong and Nathaniel Curran performed a study where they asked participants to evaluate artwork using specific criteria. They found that when participants used a specific criterion for evaluation, knowing that a piece of art was created by AI did not influence their judgment of its artistic value (Hong and Curran 2019). Although the differences between AI-created and human-created artwork may not be immediately apparent to most viewers, applying a defined criteria for evaluation affords a path for both evaluation and differentiation.

Because art evaluation can feel amorphous and ill-defined to many, Hong and Curran leveraged Schema Theory for their study. A schema is an organized unit of knowledge for an event, using previous experiences to guide a current understanding (Pankin 2013). Schemata are "an active processing data structure that organizes memory and guides perception, performance, and thought" (Norman and Rumelhart 1981, 238). Schemata serve as our cognitive compass, providing us with the opportunity to make informed decisions from previously acquired knowledge. By activating these schemata based on past experiences, we can better interpret new information and form connections between present situations and previous ones. Schemata serves as a heuristic in this regard and may be triggered by bias or stereotype; ultimately demonstrating how our past informs present decision making. Cultural information is also stored in schemas, and they can focus attention on specific aspects as a way of accommodating or rejecting aspects that do not conform. By implementing Schema Theory to evaluate the perceptions and biases of participants, Hong and Curran developed a framework for understanding their participant's evaluation of art—including art concepts, likes, and dislikes, and where they viewed art. To that end, they note that humans carry schemata that include biases about artificial intelligence and the creativity of AI (Hong and Curran 2019, 58:4).

The answer to the question, "What is good?" is nearly impossible to answer because it depends on whom you ask. Biases, knowledge, and personal experiences will frame each individual's response. The task of finding an empirical answer may appear to be a fool's errand, an exercise in defining taste. However, when the question is posed within an evaluative framework it becomes a relevant measurement of an artifact's effectiveness at communicating an idea, emotion, or information. Questions around quality and effectiveness inevitably surface when evaluating the output of any generative model, not just text, image, and sound, but also stories. When video narratives are generated by two alternate constructs, human-created and AI-created, assessment of the work's quality and relevance will likely depend not only on the structure of the story but also on the audience's bias. Cheng's work shows that visuals activate schema (Cheng 2022), which means the inherently visual construct of video storytelling will almost certainly trigger evaluations through an individual's lens of expectations and emotions.

1.5 Resistance to AI in everyday life

Earlier, I discussed how professional photographers avoided using autofocus due to its early unreliability. A second school of thinking with resistant users was that shooting photos with autofocus indicated that a photographer was not very good, that they were not concerned about the quality of their work (Edwards 2022). There was, and still is, a vocal group of photographers who believe a true artist does not rely on these cheats. They argue that leveraging this sort of assistance during creation removes the craftsperson from the craft. The size of this "manual focus only" cohort is shrinking as new iterations of the technology are introduced, and as the “never-autofocus” practitioners retire from the profession. The latest generation of photographers have learned that to thrive as a professional and accomplish better work, they need to adapt to new technology. As Mackubin Owens described these types of technological transitions, “If you don’t like change, you’re going to like irrelevance even less” (Owens 2001).

For a multitude of reasons there is pronounced resistance to incorporating AI into the creative process. However, resistance to change is not uncommon when new technology is introduced. Goldie "Red" Burns, founder of NYU’s Interactive Telecommunications program (ITP), uses an example from history to frame what she calls the danger of an “illusion of knowledge” (Maeda 2022, 44:35). She presents a fable in which, in 1895, the Lumiere Brothers projected a speeding train onto the wall of a café in France. While the audience likely knew intellectually that it was not an actual train, they had never seen a life-size, moving projected image before and, intuitively, ran out of the café. She uses this parable to explain a common response to new situations: we interpret them through the lens of our current knowledge. In the Parisian patron’s situation, they knew what a train looked like, and their illusion of knowledge told them they needed to get out of the way. The projection of the train was a product of new technology, a product they could only judge from their frame of reference, a product they did not understand.

Many of the fears about AI focus on the potential risks of releasing sentient machines into the world. A fear that development and deployment of AI robots will malfunction or turn against us (Sandler 2014). There are also concerns that the consequences of creative collaboration will result in the artist losing creative ownership of their work (Bruni 2022). One of the most vocal concerns revolves around the potential impact to workers when a creative system is significantly faster at generating creative output, and the cost of the work drops to near zero. If an AI-driven creative solution can reduce the cost of generating art, such as graphics for a website, what are the implications for the people who currently perform the work? Will one person be asked to perform the work previously done by 20 (Abram 2022)? For creatives working in a profession where jobs are often challenging to find, the fear of being replaced by a bot is real.

1.6 Why the rush to AI?

The fear of job loss by creative workers is not unfounded. Companies worldwide are running headlong to embrace AI and incorporate it into their portfolios. A common corporate rationale is this technology has the potential to reap significant financial rewards with staggering projected revenue. At the beginning of 2023 Satya Nadella, Microsoft's chief executive, expressed a belief that within the next 3 years as much as 10 percent of all data could be A.I.-generated, leading to as much as $7 billion dollars in revenue for the company (Metz and Weise 2023). In a 2022 investment report, Ark Investments projected that AI software companies will eventually generate $14 trillion dollars in annual revenue, with a cash flow yield of roughly 30%. The result of this skyrocketing revenue stream means AI software could create more than $80 trillion dollars in enterprise value, more than six times the $13 trillion in enterprise value created by the first 25 years of the Internet (Summerlin 2022). The figure to the right (*figure 3)* represents the internet's $13 trillion valuation (the green bar on the left) in relation to the potential $87 trillion dollars in revenue from implementations of AI software (the purple, blue and black bar on the right). Forecasters expect this valuation will be generated in less than half the time it took to grow Internet Era mega revenue producers like Meta, Google and Amazon (Thompson 2022).

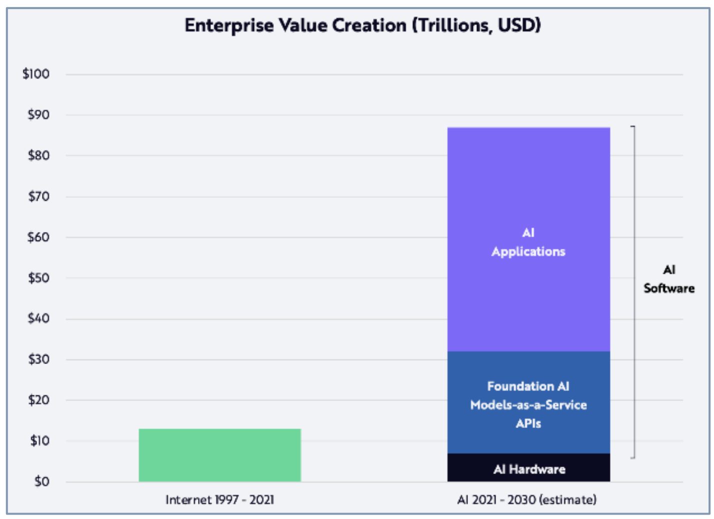


Figure 3 Enterprise Value Creation - Life Architect from: "Transformative AI vs AGI vs Singularity" (Thompson 2022).

The incorporation of AI into tools and applications has already penetrated many sectors of the economy. These applications are seen as a way to amplify market positions in areas where corporations are already a strong presence — production, banking, finance, transport, logistics, consumer technology, and electronics sectors (Carter 2018) (Kaur, et al. 2020). Companies believe this integration of AI will hedge their bets by safeguarding their innovation potential, automation, business growth, increased productivity and increased organizational efficiency (Alsheibani, Monash and Messom 2018). Aleksandra Webb writes that businesses are incorporating AI to reduce costs and grow their operations. She states that they increase profits by improving global value chains, thus increasing the financial gains of both local and global markets (Webb 2022). Companies also project benefits from offering a range of options to consumers, who increasingly incorporate this technology into their day, reaching for digital products that organize their lives (ibid. 2022).

The perception of both human efficiency and cost efficacy, getting more while spending less, is a refrain cited repeatedly in financial reports touting the benefits of AI. During the 20th Century, it is argued that science and technology solutions improved human workplace performance and saved on time of labor (Webb 2020). This workplace productivity is framed by both technological innovation and digitally mediated ways of working (Gal, et al. 2019). Digitalization, like the internet, was an essential driver of this growth, changing the way we work, learn, and communicate. Analysis of the transformation to a digitized workplace described contemporary business models as "Knowledge" and "Creative" economies (Florida 2014).

Financial analysts project that this new wave of AI-driven technological innovation will be of even greater benefit to workers, exceeding the previous 15 years of digital transformation. One PricewaterhouseCoopers study projects a future highlighting new opportunities associated with better pay, working conditions, and higher social status (Berriman and Hawksworth 2017). They argue that these sorts of sunny projections of ample employment are replacing, "the old fear of digitalization and automation being solely responsible for loss of jobs, or jobs being replaced by robots."

These projections of worker wealth and ease appear to be in direct opposition to the fears expressed by creatives. They are replacing previous fears of losing one's job to automation not with predictions of hope and productivity, but with fears of AI-driven replacement. The jobs created in the digital economy, often referred to as "Knowledge Workers", seem to be the most vulnerable to AI automation. Generative AI, for example Chat GPT and Stable Diffusion, is squarely focused on the millions of knowledge and creative workers. While research indicates that AI will make these workers at least 20% more efficient and/or creative by increasing efficiency and capability (Huang and Grady 2022), it is also worth noting that none of the reports mentioned above mention the anticipated job loss when workers are replaced by AI. There is no indication of the costs associated with employee retraining, relocation, or the personal cost of displacement when an unemployed worker is unqualified for emerging opportunities (Burley and Eisikovits 2022). These projections of technology ushering in a better tomorrow, with little regard by corporations for the potential collateral damage to workers, are not a new phenomenon, and is a topic that will be explored in detail in the next chapter.

1.7 "Turing Test"

As the quality of AI output is refined and improved, there is a likelihood that the content generated by algorithms will be regarded as "better than humans." The evaluation of computational creativity's results is a human v. machine comparison, a constant appraisal of perceived value compared to human output (Arriagada and Arriagada-Bruneau 2022). The work of Cheng, Hong, and Curran, described in section 1.4, provides examples of this type of evaluation (Hong and Curran 2019) (Cheng 2022). This process of testing a viewer's response and then weighing the perceived value of the AI-generated media is a contemporary variation of the "Turing Test" (Wellner 2021).

In 1950, English mathematician and computer pioneer Alan Turing presented a test of machine intelligence, offered in the guise of a parlor activity, called the "Imitation Game" (Turing 1950). The game involves three participants: a man, a woman, and a third participant of either sex who acts as an interrogator. The goal of the game is for the interrogator to identify which player is the man and which is the woman, based on questions posed by the interrogator to the other two players, each with opposing goals in the game. One player acts as a foil, attempting to trick the interrogator into arriving at an incorrect conclusion, while the other assists the interrogator in arriving at the correct conclusion. Turing used his game as a framework for posing a deeper question about the power of computers. The game was part of Turing's seminal article, "Computing Machinery and Intelligence", in which he outlined his criteria for considering whether a machine can demonstrate intelligence. He famously began his article with the question: "Can machines think?" (ibid. 1950, 433). To explore this question, he posited potential outcomes when player "A," the antagonist, is played by a computer. Can the machine outsmart the human? The purpose of the game shifts from an exploration of our assumptions about gender and intellect to a theoretical boundary where machines are equal to humans, at least performatively. The Imitation Game is just one small part of Turing's work in developing intelligent machines, a topic that will be discussed in greater detail in Chapter Three as part of an exploration of Artificial Intelligence's arc of development. However, the "Turing Test" has become a cultural touchstone when discussing the potential of computing.

Today, the term "Turing Test" is typically used as a shorthand measure of whether machines can convincingly imitate human behavior and interaction, to the point where a participant cannot distinguish between the computer and a real person (Hong and Curran 2019). In many ways, we perform a "Turing Test" in our heads whenever we evaluate the merit and creativity of AI-generated artifacts. What happens then, when we determine that a piece of media has successfully passed the test and even surpassed it? In "The Ethics of Artificial Intelligence" (2014), Nick Bostrom and Eliezer Yudkowsky argue that once AI surpasses human performance in a particular task, the human skills required for that task will lose value and will no longer be considered a representation of "intelligence". The impact of this shift could conceivably force us to reconceptualize our understanding of creativity's importance when assessing AI-generated media that passes the "Turing Test". Particularly if we find it more appealing than human-created art.

1.8 AI in the workplace and the "Tipping Point"

If generative AI is well on the way to becoming faster, cheaper, and in some cases better than human creations, what are the potential implications for both creators and consumers? Industries that currently require human-created original work, such as advertising, architecture, motion graphics, game development, product design, customer service, law, and marketing, to name but a few, are being reinvented. Some jobs may evolve into hybrid collaborative creativity between machine and creator. Many workers are likely to be replaced outright. Generative AI will certainly unlock better, faster, and cheaper creation across a wide range of end markets (Huang and Grady 2022), but the result will be a total reinvention of what it means to work.

There is a belief among many of my western type-A overachiever peers that one's career defines you. Dubbed "workism", these folks propel themselves through life operating with an assumption that work is not just an economic necessity but life's primary source of identity and meaning (Roose 2022). With AI seeping into the workplace, hustling to prove oneself and get ahead is probably counterproductive. It is impossible to outwork an algorithm, it would be like trying to outrun electricity. Consequently, the entire cultural construct of work hard, work fast, and get ahead will be turned on its head. Rather than trying to win at the algorithm's zero-sum game, those that survive will need to find ways to leverage the algorithm's power and amplify the unique advantages of being human. But how?

Computer scientist and author Brian Christian believes that we, "are at the beginning of a broad societal transformation" (Chen 2022, 1) OpenAI's head of policy, Jack Clark, frames the development of tools like ChatGPT and DALL·E as windows into possibilities and a primer to prepare the world for what will soon be mainstream technology (Hern 2019). He believes that "the rules by which you can control technology have fundamentally changed," because of the economies of scale, and a corresponding distribution of AI with few regulatory guardrails (Ibid. 2019, 3). There are underlying disagreements within the development community about how this technology should be deployed. Some have raised concerns that AI will mirror the unchecked development of the digital revolution, driven by the utopian dreams of founders such as Facebook's Mark Zuckerberg, Apple's Steve Jobs, and Twitter's Jack Dorsey, placing powerful new technologies on billions of devices with little to no consideration for potential harm (Roose 2022). Those advocating restraint voice a belief that the release and implementation of the technology should be measured. Others, like Emad Mostaque, founder and chief executive of the AI image generator Stability AI, argue that radical freedom is necessary to keep AI untethered from corporate influence (ibid. 2022).

This disagreement over who controls AI is rooted in differing viewpoints about likely outcomes when AI algorithms are integrated into society. If the disagreement sounds familiar, it may be because we traveled through a similar technical/cultural wilderness with the deployment of social media. Most people experience the impact of these social coding technologies daily, whose societal implications go well beyond the digital architectures of the platforms. Google Search, Facebook, Twitter, TikTok, YouTube, and others have reshaped and redefined both micro and macro cultures worldwide. Jose van Dijck points out being social was not simply "rendered technological" when it moved to an online space, "coded structures profoundly altered the nature of our connections, creations, and interactions" (2013). The algorithms of these platforms intervened in the shape of our cultural encounters, manipulating the circulation and promotion of popular culture (Beer 2013). These channels became conduits of culture, distributing packets of media that reflect the interests and vision of individual creators. A creator shapes those packets during production to reflect the cultural norms and expectations of the intended audience (Shifman 2014), hoping the algorithm will reward them with a connection to people who share the common culture (Chayka 2020). This distribution of culture via algorithmic channels became "a ritualized collocation of different people on the same mental map, sharing or engaged with popular ontologies of representation" (van Dijck 2013). These packets of media (be they photos on Instagram, personal events on Facebook, tweets on Twitter, or video on YouTube) are human-created cultural objects, symbols that can be interpreted as parts of a code (Parsons 1970). The algorithms became our personalized virtual bookshelf, filled with algorithmically driven cultural packets, constantly shuffling and moving on the shelves, driven by our engagement, demographics, and geography. Through our attention, measured and tracked by the algorithms, we empowered a system of cultural filtration to be the collector, curator, and arbiter of culture, undermining human-driven cultural connections (Chayka 2020).

If research into the effects of social media algorithms by van Dijck, Beer, Chayka, and others demonstrates a negative impact on society's cultural connections, should we expect AI's algorithms to have any less of a deleterious effect? The answer to that question will likely be driven by AI's traction with users. If the adoption rate of artificial intelligence is low, and it is largely unused and ignored, the handwringing projections of a dystopian robot culture in the popular media may be little more than a tempest in a teapot. However, if users are testing the technology, embracing its potential, and sharing the output with their communities, does that indicate that they see value in these tools?

The novelty and power of recent innovations like GPT and Stable Diffusion are clearly attractive to early adopters, as user counts indicate that utilization is coming quickly. Within one week of its November 30 release Open AI's ChatGPT saw one million users, with an expectation it will cross the 1 billion user threshold within a year (Ruby 2022). Stability AI, the company leading the development of Stable Diffusion, announced that as of October 2022 over 10 million daily users of the platform had created 170 million images (Wiggers 2022). These early adopters are creating a groundswell of interest, and one result is a corresponding tsunami of press coverage. A simple search in Google for "AI News" results in over 2.5 billion results (Google 2023), with countless news articles, in every media channel, exploring the benefits and shortcomings of not only these two applications, but all things AI. It is evident that the topic of Artificial Intelligence has moved from research papers to the newspapers, but will it manifest self-sustaining adoption rates, propelling AI to what Everett Rogers referred to as the "tipping point" (E. Rogers 2003)?

Everett Rogers's *Diffusion of Innovations* (2003) and his identification of the tipping point have morphed from an academic reference to shorthand used to describe an invention's popularity in popular media. The notion of the tipping point – when a trend reaches critical mass, accelerating its spread in a domino effect and consequentially effecting change in an entire social system – finds its conceptual foundation within diffusion theory. This is a set of generalizations regarding the typical spread of innovations, for good or bad, within a social system. The focus of the theory is on defining how a potential technology adopter decides to use something new. A user's evaluation of an innovation's worth involves a cost-benefit analysis, often shrouded in personal consequential uncertainty. Rogers writes that people are more likely to adopt new ideas if they determine that the overall utility can be improved with a corresponding shift in thinking because of adoption, suggesting there must be some sort of relative benefit over previously held beliefs (ibid. 2003). A user is constantly measuring the potential costs and liabilities as they consider adoption: How will they know with certainty there are benefits? Will it be disruptive to other facets of their daily life? Is it compatible with existing habits and values? How will others in my community perceive me? Does it really work? A user will adopt an innovation if they believe that it will enhance their utility with a minimum of liability.

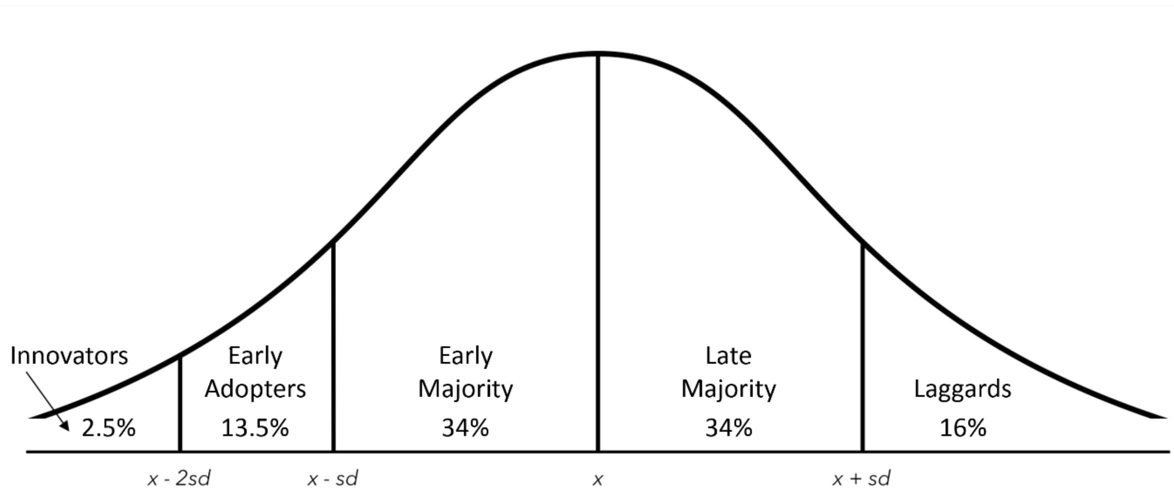


Figure 4: Relationship between types of adopters classified by innovativeness and their location on the adoption curve. Source: Everett M. Rodgers, Diffusion of Innovations, 5th ed. (New Yor Free Press, 2003), p.281

Rodgers plotted the adoption rates of users across a simple bell curve, classifying them into discrete categories: Innovators, Early Adopters, Early Majority, Late Majority and Laggards (ibid. 2023, 262). His curve identifies the degree to which individuals are comfortable with innovation. The less anxiety a potential user feels over the potential disruption resulting from adoption the sooner they are likely to embrace a change. Each member of a social system faces his/her own innovation decision, and Rodgers wanted to identify the process by which an innovation is communicated through certain channels over time among the members of a social system (ibid. 2023, 162).

An important aspect of diffusion theory is Rogers' belief that a decision to adopt an innovation depends on the innovation-decisions of other members in the system. The choices made by others have a direct impact on the cost/benefit calculus of a user. Some members are risk-averse and postpone a decision until they gather more information, while those who are more venturesome, the Innovators, enjoy being on the cutting edge and give something "a try" because of the anticipated excitement over perceived benefits (ibid. 2023, 263). Enthusiasm shared by these Innovators about the implementation of the technology, confirming the benefits of their choice, becomes a catalyst for others, casting the Innovators in a trusted advisor's light. As more members of the social system follow suit, the adoption rate grows, potentially reaching a fabled "tipping point" where the rate of adoption increases exponentially.

While this model is an effective way to understand how social systems adopt new behaviors and technologies, such as electric cars, fashion, slang, and social media platforms, most innovations fail to achieve the sort of tipping point level of adoption that results in hundreds of millions of users. When the number of people using a technology is small and implementation is limited, a Rogers curve is still applicable for charting their growth, but the slow adoption rate is unlikely to generate a substantial user base, nor will it reach widespread adoption.

If we chart the adoption rate for artificial intelligence as a whole, I propose that we are still in the Innovators segment of the Rogers Curve. While individual test beds with high adoption rates, like ChatGPT, are novel and interesting to the general public, the implementation of AI proposed by global companies and developers alike has barely been integrated into devices, enterprise software, automobiles, distribution centers, supply chains, financial transactions, security, and more. As discussed in section 1.6, global companies see a promising future with AI and are proving their commitment with capital by investing billions for corporate-financed development. Therefore, perhaps a better application of the Rogers curve would be charting the global macro implementation of AI as companies, large and small, begin to incorporate AI into their business model. This is a nascent technology gestating in its early phases of ideation. We are only now seeing glimpses of how these organizations plan to incorporate AI into all facets of our day. The time when we can assess the consequences of this implementation – both beneficial and detrimental – is still on the horizon.

1.9 Societal consequence of new tools

The mechanics of diffusion provide insights when considering the mechanisms of innovation in social systems. With identifiable corporate investment and popular media attention, it appears that the first tendrils of AI innovation have arrived, and the diffusion of large-scale implementation is at our doorstep. While the larger cultural impact of this diffusion is still a guessing game, the probable effect on content creators is more readily identifiable. There is hope that with careful thought and consideration, we will manage to control the impact of AI without causing harm to ourselves or others (Chen 2022). Although recent history shows that there is a stark gap between the promises of new technology and the real-world experience of those affected (Roose 2022).

One fear expressed by those concerned about creative worker job loss is the sheer scale of AI's deployment. They paint a bleak picture where countless artists struggle because of lost work (Ramesh 2022). On the other side of the argument, proponents envision meaningful benefits for corporations when AI generated media is indistinguishable from human creations (Thompson 2022), a point of view that appears to celebrate technology over the value of human creators. While each side holds differing perspectives about AI's benefits and liabilities, both agree that its integration into creative industries is a fait accompli.

I have personally reached the conclusion that AI's march of progress is ushering in yet another instance where I will need to master a new technology. This decision is motivated by my desire to sustain occupational relevance. Many of my previous adaptations have required finding paths for retraining, and strategies for retaining, my creative voice. Cultivating a mindset of reinvention has not been a noble exercise on my part; instead, it is always a question of survival in a craft that defines who I am and forms my concept of self. These adaptations were figuratively "adapt or die" decisions. This raises a personal question for me: if machines can now, or soon will, generate creative work that is "indistinguishable from human creations," where do I fit into this creative milieu?

. The questions I have posed here are shaped by our definition of creativity. Machines are demonstrating the ability to come up with new, novel, and valuable ideas in surprising and unfamiliar ways. The fact that AI-generated artifacts are nearly indistinguishable from human work, often perceived as equal or superior to human effort, is forcing us to examine the emotional and social value of creativity. These synthetic artifacts are appreciated because viewers ascribe messages and emotions to the work, despite the absence of a human originator's perspective. Clearly, creativity in all forms can be viewed as a mystery, a puzzle, and a paradox. It is an expression of the unfamiliar (Boden 2004). So, if creativity is "part of what makes us human" (Sawyer 2012, 3) how does our perception of an artifact's worth, and a cultural value of human contribution, shape the future employability of a creative worker? How can creatives differentiate their work from work generated by robots?

AI is a powerful tool that will almost certainly take jobs away from artists. There is demonstrable ease with these tools in storyboarding an idea, developing concept art, or bringing new ideas to life. When a machine can essentially accomplish the same task and bring the cost down to near zero it becomes impossible for a creative craftsperson to compete (Ramesh 2022). Consequently, creative jobs will begin to disappear, not unlike the experiences of portrait painters in the 1800s after the introduction of photography (Sutton 2022). Current applications like DALL·E 2 excel at creating photorealistic images like food photography or the type of images used in corporate brochures and websites. Researcher Eliza Strickland points out that AI generated illustrations "wouldn't seem out of place on a magazine cover" (Strickland 2022). Motion designer and educator Yannick Theunissen believes, "AI can do so much that many of our jobs are probably at risk" (Theunissen 2022, 01:11).

Our appreciation of a specific creative object reflects both the aesthetic and emotional facets of human intelligence. We measure the interplay between the artist, society, and the environment (Corazza 2016). Artificial Intelligence is tilting this calculus away from human-generated voices to machine voices, with AI bots generating artifacts that synthesize fragments from the history of humanity. For the last decade, the internet and globalization have created a landscape where creative professionals are already competing against hundreds of thousands of other creative workers for opportunities (Edwards 2022). In the massive marketplace of Fiverr, UpWork, and other freelance platforms, success is often a reflection of a creator's unique cognition and hustle. AI is likely to turn this world upside down. One consequence of AI in the creative marketplace is the probability that creative workers will need to shift their focus. For many, their efforts may require less internal creative exploration and more time spent operating at scale, developing strategies for integrating AI into their craft.

For some, the key to survival will be discovering ways to differentiate their expressions of human creativity from the latent space of an AI model hosted on a distant neural network. Those who not only survive but thrive are likely to reach beyond the perceived limits of AI's homogenized point of view, pushing aside the banal in search of a creative voice that is one part personal and one part machine. The creative content we see going forward will likely stratify into three categories: work created by a sociotechnical ensemble generating AI content with the creative-nutritional value of junk food, bespoke specialty content that is a unique reflection of the human creator, and hybrid models where AI becomes an inspirational tool blending machine creativity with the artist's vision. The question is, what will be the ultimate percentage of each stratum? How many opportunities will there be for those presenting their personal point of view?

We are seeing that the function of artificial intelligence in the creative arena is more than a tool, it is an actor with artistic and creative agency. These technological advancements are certainly a leap forward, carving out new arenas for creativity. However, the dilemma going forward is how to avoid a precipice where creatives are forced to choose between the job of managing AI or unemployment. It would be an unwinnable conundrum for creatives. Perhaps we can find solace in the belief that real intelligence, human intelligence, has some advantages.

Chapter 2

Industrial Revolutions: From Steam Power to AI

2.1 Wrestling with change

Part of the human experience is wrestling with change. We fear the unknown and the uncontrollable, events driven by entropy and chaos. For some people, that anxiety is paralyzing. For others, it elicits excitement and visions of imagined outcomes. For most of us though, it is an ever-shifting tide of fear and hope. When change appears inevitable, we somehow summon the courage to grit our teeth and step carefully into the unknown.

As someone working in a creative field, I have never considered the possibility that someday a robot might take my job. That anxiety was never on my list of worries. I've always considered video production a messy process more analogous to sausage production than calculus. Crafting a story has always felt complex, nuanced, and intrinsically human. The concept of an algorithm breaking my craft into patterned steps and performing the task was beyond my realm of possibility. Yet with the rapid advances in Artificial Intelligence, I catch myself wondering if my job will go the way of lamplighters and town criers. I now feel the need to face the question: "Is my obsolescence on the horizon?"

The introduction of AI into the workplace is very likely ushering in a new era, an industrial revolution as consequential as the 18th century introduction of steam power and the dawn of the first Industrial Age. Journalist Jon Talton writes that the potential upheaval associated with Artificial Intelligence could forever alter life and society (Talton 2022). Physicist Stephen Hawking warned that AI could mimic human intelligence, surpass it by becoming independent, and ultimately it “could spell the end of the human race” (Cellan-Jones 2014).

Even if we do not realize Hawking's vision of living in a dystopian world run by sentient machines, it is almost certain that AI will increase efficiency and lower the cost of tasks currently performed by humans. As Eray Eliaçik points out, the result of these changes will have consequences, with the net result of AI adoption implying fewer jobs, resulting in personal and global economic complications (Eliaçik 2022). These complications, manifested as social and cultural events mirroring similar shifts in prior industrial upheavals, will shape how we think, how we create, how we preserve our ideas, and how we see ourselves in society. By examining these previous transformations, we can draw parallels that illuminate common outcomes in the struggle between potential and fear, rich and poor, and the known and unknown. These parallels offer insights into how innovators often blindly race towards invention, how craftspeople and workers are rarely the first, second, or even third consideration when driving change, and how outcomes are usually unimagined. The recent explosion of discussion about AI's potential and our potential peril is forcing us to recognize that this revolutionary change is here, now.

The implementation of Generative AI is not theoretical for many media creatives, as these tools are already making their way into production paths (Roose 2022). For graphic designers, applications such as Midjourney can churn out polished, detailed images in just minutes, at a cost of only pennies per graphic (Crabapple 2022). Stable Diffusion is an open-source image model, meaning it is available for free. Anyone can view the code, download it, modify it, and run the software on their personal computer (Stability AI 2022). Like Stable Diffusion, Open AI’s DALL·E 2 is a generative model that turns text descriptions into vividly detailed, high-resolution images (Samuel 2022). Adobe has released their version of a text-to-image model and incorporated it inside applications like Photoshop and Lightroom (Belsky 2022). Content creators are generating synthetic on-camera video presenters with tools like Rephrase.ai and D-ID. With D-ID’s Creative Reality Studio, a user uploads a “head shot” photograph for training the AI, which then generates a photorealistic avatar. This synthetic likeness is then combined with an AI generated voice to produce an entirely synthetic video (D-ID 2022). In 2016 an AI named “Benjamin” wrote the script for the science fiction film, *Sunspring* (2016). That year the film appeared at the Sci-Fi London Film Festival, beating out hundreds of other human-scripted films for its place in the festival. Researcher Deniz Kurt makes the point that this selection was important because it implies the AI-generated story successfully represented human emotions within the narrative (Kurt 2018, 33). For music generation, creators can prompt Soundraw to compose fully orchestrated and mixed songs based on parameters like style, tempo, length, and instrumentation. Within seconds it presents 10 options for the user's consideration. They can choose the one that most closely matches their needs and finetune the length, the arrangement, and even the energy level of their composition (Soundraw 2020). The speed of media generation, and diversity of options, can be unnerving to users. Yannick Theunissen states, "Personally, this is one of the scariest parts of creative AI, how it can generate realistic stuff in moments out of nothing more than a text prompt. Do my skills even matter?" (Theunissen 2022, 03:34)

Skilled workers expressing a fear of lost occupational relevance and ultimate displacement is a refrain heard with every technological leap, echoed in the academic writing during times of technological change. During the age of mass production economist John Maynard Keynes predicted that technology would drive widespread unemployment "due to our discovery of means of economizing the use of labor outrunning the pace at which we can find new uses for labor" (Keynes 1931). During the digital revolution economic and social theorist Jeremy Rifkin argued that "Today, all ... sectors of the economy ... are experiencing technological displacement, forcing millions onto the unemployment rolls" (Rifkin 1995). In the economic slowdown of 2016 scholars pointed to computer-controlled equipment as a potential driver of jobless growth (Frey and Osborne 2017). Each of these describe scenarios where job-loss is tied to task automation, even when the impact on the worker's status is not instantaneous. Technologist Brian Merchant writes that contemporary "automation does not appear to immediately and directly send workers packing en masse" (Merchant 2019). Instead, he believes that the impact of automation in the workplace surfaces gradually through pay cuts, turnover, and unfilled openings.

Yet optimists counter those claims with the argument that hundreds of years of evidence illuminate different long-term outcomes. In *The Fourth Age: Smart Robots, Conscious Computers, and the Future of Humanity* (2018), author Byron Reese states that while technology destroys some jobs, it always creates new opportunities, raising the standard of living for workers. He concedes that the steam era displaced farmers and cottage workers, but it also created millions of jobs in new factories. Factories that were producing cheaper and readily available consumer goods. Reese believes this pattern of initial displacement, followed by job creation, has been repeated throughout human history (ibid. 2018). He writes that when a technological innovation results in the loss of one craft, it ultimately creates a multitude of new job types. Electricity saw the end of lamplighters, icemen, and Victrola salesmen, but it created the entirely new industries of electric lights, home refrigeration and radio. These new manufacturers were creating previously undreamt jobs as they mass produced the electrified gadgets that people designed, manufactured, sold, and repaired.

Pessimists and optimists appear to see the outcome of technological revolutions from dissimilar perspectives, but I believe they are both right. As we will see, short-term innovations are often the catalyst for jobs lost, worker displacement and social upheaval. Over time though, these same innovations result in the creation of entirely new industries, employment opportunities and social realignment.

2.2 Efficacy and cost cutting

The fulcrum of change, the intersection of old and new, is an uncomfortable place. Our discomfort is amplified when the technology at hand is immature, retarding user adoption, clouding potential outcomes, and making decisions problematic. Do you adopt, fight, or wait? That propensity to feel unsettled with the unknown is one way to view this moment in time as we make personal decisions about the implementation of AI in our lives. Clearly global companies have made their choices known with large capital investments. What is still undermined is if corporate customers will buy these new products, and if individuals will decide the benefits outweigh the risks. For individuals, is likely that our acceptance of AI will be shaped by a series of factors.

The first is we live in a culture of connectivity brought on by the digital revolution. Our relationships to the culture we consume has transformed from physical to digital, we have radically altered our patterns and volume of consumption (Chayka 2020). Author José van Dijck writes that this culture of connectivity has evolved as part of a longer historical transformation where the boundaries between private, corporate, and public domains have been reset (J. van Dijck 2013). These blurred boundaries have resulted in a dynamic where we are courted across multiple digital platforms streaming algorithmically chosen stories, brand identity messages and a constant drumbeat of consumerism (Jenkins 2007). The messages we hear, the messages that shape our choices and beliefs, the messages that drive the adoption of innovation, are fed by a culture of connectivity that is completely integrated by these media infrastructures (Beer 2013). I would argue that contemporary media attention on the potential benefits, and “magical” computational creativity generated by AI, has created in society a predisposed interest in AI, lowering the barriers to adoption and ultimately creating a foundation of acceptance.

The second factor is we are already seeing the effects of workers being replaced by AI automation. There is an ever-growing list of jobs where it has been determined that a computer can perform comparably, or better, than a human. From fast food preparation to insurance claim processing to taxi driver (Waller 2017). In the United States retail stores like Amazon Go and Whole Foods have implemented the Amazon One payment system, where a customer's selections are tracked as they roam the store and payment is completed with a biometric wave of the hand (A. D. Thompson 2022). There are no cashiers involved in the shopping experience, the customer is entirely independent, with Amazon calling it a 'just walk out' shopping experience (Tillman 2022). Amazon has even experimented with a select number of Amazon Go stores by essentially replacing the entire staff with AI, implementing a completely automated system with no visible employees.

The third contributing factor is western culture's embrace of capitalistic development and exchange structures. Our advanced capitalism is an economic engine that promotes the quantitative proliferation of consumer goods, producing deterritorialized differences through the promotion of commodities. Our social spaces are saturated by fast-changing commodities that generate what Brian Massumi refers to as the "contradictory temporality of commodity fetishism" (Massumi 1992). We have developed an appetite for things that never fully appease, seeking the "new" in a system that Rosi Braidotti described as "addictive" (Braidotti 2019). This journey on the treadmill of consumerism predisposes potential users to find the shiny newness of AI "attractive". The novelty of media being created inside a magic box promises to feed a hunger we are unable to appease.

Big companies are hoping to feed the appetite of all types of users by developing AI applications in the name of productivity. Applications that promise to improve a company's bottom line or attract consumers. Financial analyst Will Summerlin believes that AI software will boost the productivity of the average knowledge worker by 140%, adding approximately $50,000 in value per worker, or $56 trillion globally (*figure 6*) (Summerlin 2022). Paul R. Daugherty and H. James Wilson, two executives with the consulting firm Accenture, write that human-AI collaborations will be a cornerstone of the twenty-first

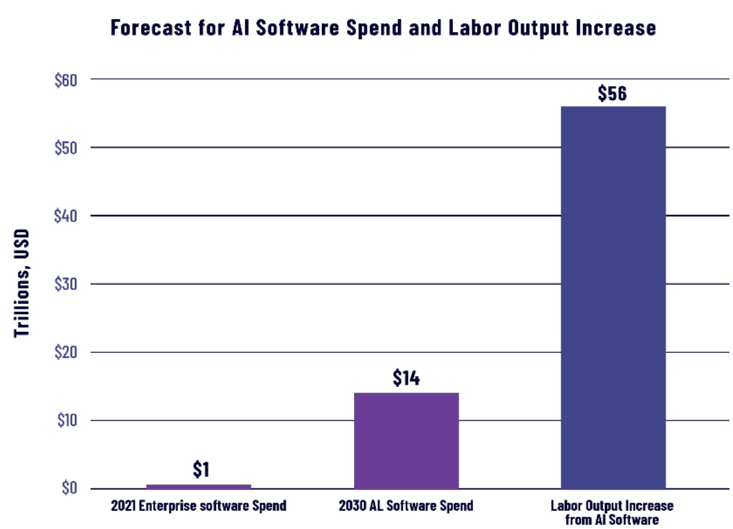


Figure 6 Adapted from: ARK Investment Management, 2022.

century economy. “AI systems are not wholesale replacing us. Rather, they are amplifying our skills and collaborating with us to achieve productivity gains that have previously not been possible,” (Daugherty and Wilson 2018) Companies like Mattel, a $6 billion per year toy company, have already implemented the human-AI creative hybrid model as they develop new toy concepts. In 2022, they iterated designs for their Hot Wheels series of cars by using DALL·E to generate alternative designs and colors of their latest toys (Roach 2022). Microsoft’s integration of AI into their design application, Designer, appears to position the technology as just another tool for creating lovely images to share with friends, colleagues, clients and customers (Hachman 2022). At a 2023 press conference, Microsoft’s Panos Panay confided that “AI is going to reinvent how you do everything on Windows, quite literally” (Hollister 2023). There are countless other examples of profit projections, anticipated productivity gains and promised consumer benefits in the rush to capitalize on AI. In the entropy of development there is little transparency about the process of assembling the parts, with innumerable assurances about a rosy outcome.

One of the engines of this capitalistic system is the appropriation of work done by others with no compensation, in essence monetizing the work of the crowd (Rosenbaum 2013). Similar to how Google Translate is based on aggregating the work of thousands of individuals who translated an enormous body of phrases to train the system’s algorithm (Schuster, Johnson and Thorat 2016), the development of text-to-image generators rests on the hundreds of millions of images and graphics created by human artists. In both instances, the original creators received no compensation, and the result is a system where the developer is in essence selling a creator’s work back to themselves.

It is unlikely the translators, artisans and artists imagined the consequence of sharing their work online. The digital paths that enable our culture of connectivity have been promoted as conduits for celebrating work, business development, and educating others. As individual bits of data it was difficult for creators to assign monetary value to the tiny outputs they freely shared. Each translated sentence, each drawing, each fantastical illustration was little more than a droplet in a sea of creativity. The value was in the aggregate of work. Only now, with the realization of AI’s potential impact on their livelihood, do they have a conceptual basis for questioning the outcome of this creative appropriation.

The difficulty of imagining consequence continues to be a challenge for many creators. Most do not have a clear understanding of either the computational process nor ease of use with diffusion applications like Midjourney and DALL·E 2 (Theunissen 2022). When a user with no technical skill nor artistic training can generate media, the law of supply and demand tells us that the corresponding value of a professional’s output will diminish. Many creators have taken a fundamental stance described in chapter one that while AI performs impressive tasks, it will never replace humans because it is incapable of imagination, creativity, or a human perspective (Brownlee 2022). It lacks the humanity of human-generated work. Even when the work generated by a computational creative model is objectively similar to those generated by humans (Schroer and Glover 2022), the work is dismissed out of hand because of an innate conviction that art exclusively emanates from human observations and efforts (Cheng 2022).

Initially, Silicon Valley futurists anticipated AI's impact on the workforce would be with traditional blue-collar workers like cashiers, warehouse workers and truck drivers (Roose 2022). Just like factory tasks were mechanized on assembly lines during the era of mass production, these repetitive manual tasks could be automated via AI. They believed workers in creative fields like art, entertainment and media would be safe. What they did not anticipate was the consequence of technology reducing the cost of image creation to near zero. The likelihood of a wholesale shift to computational creativity in the graphic arts is very high as many organizations are doing everything to increase profitability by cutting costs (Abram 2022). This shift is happening in an industry where it is already difficult to find and keep work, and the job description is likely to change from graphic designer to prompt designer (Merzmensch 2022). As a result, creators are searching for paths that afford them a way to maintain their artistic credibility while also generating an income. Video producer and educator Jordy Vandeput describes the struggles with this adaptation as, "Just like Blockbuster went bankrupt because they didn't embrace streaming, this is the fear we are all battling as the technology continues to advance" (Vandeput 2022, 00:00).

Many creatives initially dismissed the idea that AI would pose a threat to their craft. Journalist and author Kevin Roose writes in his book *Futureproof* (2022) that he imagined Artificial Intelligence might facilitate the more mundane, routine tasks involved in journalism. Jobs like fact checking, data analysis and story outlining. He did not believe they would be capable of executing the parts of the job that required creativity, imagination, human interaction, and explanation, "but as I thought more about it, I started to worry that I'd been deluding myself" (Roose 2022, 63). With the rapid advances in the capabilities of AI some artists are questioning whether the technology is focused less on helping creatives and more on being an artist replacement. They see direct parallels between today's rise of AI and the technological upheaval associated with the introduction of steam power. Artist and writer Molly Crabapple likens the adoption of Generative AI to the introduction of the self-acting spinning mule, "a machine commissioned by British factory bosses in 1825 to break the power of striking textile workers" (Crabapple 2022). Video journalist Phil Edwards draws a similar parallel between the development of AI-created artwork and the fall of 19th century handmade lace production, noting that when production was mechanized the market for human-created lace disappeared (Edwards 2022).

The feeling of unease voiced by creatives has merit. Corporate investment in AI tells us that the technology will become even more pervasive. The connectedness of the digital revolution and the communities fostered by social media created cultures in which creative output was freely shared, rarely copywritten, and eventually aggregated as algorithmic training datasets. It is unfortunate that creatives appear to have inadvertently facilitated the training of the bots that are threatening their jobs in the next revolution.

2.3 Lace

Measured, recorded, collected, researched, and celebrated, lace is a textile in which the pattern is surrounded by air. The beauty of handmade lace is defined by its delicacy, the value is determined by its exclusivity. The complexity of the pattern reflects the weeks, months, even years it took to make. The intricacy of lace's individual threads connecting to create shapes woven into a patterned lattice make it a celebration of beauty, reflecting the artist's industry and attention. In her definitive anthology of lace, *A History of Hand-made Lace,* Author Emily Jackson describes wearing lace as "a thing everybody cannot have. That it proves, by the look of it, the ability of the maker" (Jackson 1900, 10).

Figure 7 Elizabeth I, 1533-1603 (the 'Armada Portrait') National Maritime Museum, Greenwich, London. (Gower 1588)

Handmade lace, or *passements*, has always been a symbol of opulence, a luxurious accessory rather than a necessity. As such, price was dictated by those who could afford it (Farrell 2007). From its introduction it has been a moniker of status, a luxurious symbol of power and wealth. Consider the 17th century portraits of royals, nobility, and landowners. They chose to follow the fashion of the day by wearing lace as a differentiator from the lower classes (Ball 1922). The Privy Expenses and inventories of England's Queen Elizabeth I's New Year's gifts display an explosion of "notices" documenting her lace collection. Jackson effusively describes her love of lace by nearly equating it to a ‘little black dress that goes with everything’:

> It is the one costly wear which never vulgarizes; jewels worn without judgment can be rendered offensive to good taste in their too apparent glitter, but lace in its comparatively quiet richness never obtrudes itself and is recognized in its true worth and beauty only by those whose superior taste has trained them to see its value (Jackson 1900, 20).

Lace production was an important economic engine in 18th century Europe, providing income opportunities for women as a rural cottage industry. It is estimated that 100,000 workers were employed crafting lace by hand at the beginning of the century (Freedgood 2003). The industry largely employed women and girls, whose finesse with "nimble and slender fingers" was required for the detailed work (Freedgood 2003, 627-628). Unlike tending livestock or working in the fields, lacemaking was not laborious and could be done at home. The craft required only a few tools and working materials, engaging women who were otherwise unable to venture into the workforce. With the low cost of materials their profit was derived from time and labor. In her book, Jackson projects a lacemaker’s feeling of satisfaction with the employ because creating lace afforded an expression of artistry and individuality where “the taste of the worker is so great that a very high value can be obtained by the humblest operator” (Jackson 1900, 19).

Lace was a product of fashion, with growing demand in the seventeenth century driven by colonial expansion and export to the British colonies of North America, India, and the Dutch South African colonies (ibid. 1900). With lacemaking employing over a hundred thousand workers, and demand outstripping supply, 18th century inventors believed they could realize substantial profit by creating machines that could manufacture lace at scale. The challenge they faced was automating a complex process that required hundreds of steps. 1768 saw Anthony Hammond adapt an existing automated weft-knitting machine, called a stocking frame, into a process for crafting a simple lace pattern (Palliser 1869). While his invention was functional, it could not produce the highly prized “pillow lace”. The next forty years saw a succession of failed attempts to automate production. Each inventor

Figure 8: Drawing from "A History of Hand-Made Lace", E. Jackson, 1900

believed he had developed a method for crafting the complex stitches of handmade lace, with multiple patents filed, ultimately resulting in disappointing production (ibid. 1869). Development of these iron behemoths required significant capital investment, with costs borne by large textile manufacturers who saw the potential to corner the lace market. In 1809 John Heathcoat invented a bobbinet machine which could make an exact copy of East Midlands laces, leading to a patent and the development of Leavers, Pusher and Curtain machines (Farrell 2007). To develop the process, he studied the hand movements of a Northamptonshire lace maker, replicating them in a roller-locker machine (Earnshaw 1994). Little did she know at the time, her contribution was likely to end her career.

The bobbinet machine's emergence signaled an era that heralded the promise of eliminating manual labor's tiresome repetition. Although the quality of the initial product was often poor, machine-made laces were considerably cheaper to produce and therefore more affordable for customers (E. Jackson 1900). The expansion of machine-made lace production created thousands of new jobs for qualified workers. The principal qualification? That they were male. In a pattern that we will see repeated with each successive technological revolution, the automation of lace production eliminated some jobs while creating new ones. For the displaced workers, the disruption would prove catastrophic as the new jobs required technical skill, physical strength, and what factory owners felt was the proper gender. The opportunities created were simply not available to the displaced workers. The industry rationalized this decision to exclusively employ men by pointing out that running the mechanical looms required considerable physical strength. With their rhythmic, mesmeric chugging, the machines were massive and were loud enough to shake a building. Viewers of the weaving process expressed surprise over the juxtaposition of the heavy metal creating spools of delicate lace (O'Riordan 2022). In a research article titled "Fine Fingers: Victorian Handmade Lace and Utopian Consumption" (2003), Elaine Freedgood offers an 1833 quote from an ardent proponent of machinery and automation, Andrew Ure. He believed that "the most perfect manufacture is that which dispenses entirely with manual labor" (Freedgood 2003, 629). By the 1860's, with the introduction of the Jacquard loom, the industry was fully steam-powered, employing 150,000 predominately male workers producing lace by machine. The Jacquard loom could imitate virtually every form of handmade lace (Wardle 1968) and reduced the physical demands on the workers. Men remained as the primary workforce, acting more as minders than operators (Freedgood 2003).

From its apogee of perfection in the 17th century, handmade lace began its decline with the introduction of automation. The impact of this shift from handmade to manufactured lace on the more than 100,000 women working from home was devastating. In the fifty years since the introduction of the bobbinet machine the art of handmade lace was dying (E. Jackson 1900), with Emily Jackson marking the year 1818 as the end of the production of "old lace" (ibid. 1900, 47). The next four decades saw handmade lace prices drop to a point where demand for the products created by these cottage workers disappeared almost entirely, and by 1860 less than ten thousand handworkers were left in England (Freedgood 2003).

Before the advent of machine-made lace these workers were contributing to their household income and increasing their family's standard of living, while also bolstering the local economy. Now they were without a job (ibid. 2003). Their options for new work opportunities were limited because even if they were to uproot the family, leave their community, and move to a lace manufacturing town, their gender meant factory jobs were closed to them. A consequence of the innovation of the bobbinet machine was nearly the entire workforce of women had been removed from the labor pool. Historian Sonya 0. Rose puts a fine point on this opportunity loss by writing that the machine-made lace industry "provides a good example of how employers structured work so that women were excluded as a potential source of skilled labor (Rose 1992).

Machine-made lace's popularity grew from its introduction. The low cost, pattern variety, and lace's historic moniker of wealth and power made it an appealing fashion element (Palliser 1869). The ubiquity of incorporating machine-made lace into everyday garments prompted occasional responses by the wealthy for handmade lace, seeking the status of owning a textile made by hand. This resulted in sporadic spikes in demand (E. Jackson 1900), but eventually the overuse of machine-made lace in ready-to-wear garments caused all types of lace to fall out of style (Freedgood 2003). At the end of the 19th century lace makers tried to reinstate lace's exclusivity by marketing their product as 'real' lace, producing new types and varieties of handmade lace (Farrell 2007). By differentiating their work as 'real' or 'handmade' they were endeavoring to foster an association of intrinsic value. That implied correlation to quality is still in use, extending the relationship from practical products like bespoke furniture, shoes, and kitchen cabinets to haute couture.

Job loss, opportunity loss and personal dislocation driven by the collapse of the handmade lace market are but a few instances of social change during the era often referred to as the Industrial Revolution. Klaus Schwab, founder and executive chairman of the World Economic Forum, asserts that this era's common identification should be delineated into four discreet revolutions defined by specific technological leaps (Schwab 2017). He identifies the eras as Steam Power (1760-1840), The Age of Science and Mass Production (1900-1940), The Digital Revolution (1960-2010), and today's Rise of Artificial Intelligence (Schwab 2017). The patterns of rapid technological development resulting in worker disenfranchisement are well documented in each of the first three eras that Schwab identifies. Therefore, I believe it is possible to examine the patterns of action and response as a model that foreshadows Artificial Intelligence's potential disruption of contemporary Western society. Historically, the disruptions could be chaotic, leading to social unrest and organized labor movements. To explore the upheaval during the era of Steam Power it is worth examining the rise and fall of the Luddites.

2.4 Luddites and Engle's Pause

At the start of the Age of Steam Power societies were largely agrarian, with an economy based on farming and small industry. Most people in the textile trade lived on modest farms and in the rural hamlets scattered across northern Europe (Schwab 2017). The production of goods like handmade lace were the very definition of 'cottage industry', as many of the lacemakers lived on subsistence farms (Freedgood 2003).

With the development of hydro-powered textile manufacturing, rural laborers began a migration to factory towns in search of job opportunities. Unskilled factory workers were viewed as cheap and plentiful. They routinely faced brutal conditions in overcrowded and unsanitary factories and were often subjected to long hours and horrendous exploitation. Laborers were paid pitifully small wages, packed into squalid boardinghouses, and abused when they failed to meet their bosses' standards (Griffin 2013). With sordid living and working conditions factory workers began to organize strikes and work stoppages, endeavoring to realize better working environments and higher wages (Cherry 2019). Tired of the workers' leverage, mill owners looked for ways to reduce the workforce with the goal of undercutting their clout (McGinnis 2020). For someone living in 1820, 1783's introduction of the steamship and 1804's introduction of the steam railroad radically transformed personal mobility, but the most profound impact to their lives occurred four decades earlier with the invention of the Watt steam engine (Scherer 1965). In 1825, a Manchester engineer named Richard Roberts iterated on the Watt engine and invented the self-acting mule, a machine that automated the incredibly complicated task of spinning yarn (Waller 2016). In his book *Iron Men (2016)*, author David Waller quotes one observer who enthused "I have stood for hours admiring the precision with which the self-actor executes its multifarious successions and reversals of movement" (ibid. 2016). Dubbed the "Iron Man", it appeared to move and think as if it were a human being. Roberts' invention was a cast iron manifestation of a worker that never stopped, never took breaks, did not call in sick, and most importantly to the mill owners, did not go on strike.

With the explosion of steam-powered manufacturing, and the resultant loss of work, craftspeople faced a future where their skills were considered irrelevant (Cherry 2019). Industrial steam-powered manufacturing upended the careers of blacksmiths, farmhands, and other manual laborers, costing tens of thousands of artisans their jobs (Lovett 1876). A rising tide of discontent fueled these workers' desire to fight the rush to industrialization with direct action against the mills and factory owners. Luddism was a direct response to both job loss and the poor working conditions in the emerging industry of textile manufacturing.

The English Parliament had outlawed unions under the Combination Act of 1799, significantly undercutting labor's ability to organize (Lovett 1876). It was amidst the stark conditions of uncertain employment, shoddy medical care, nonexistent safety standards, and food shortages that the Luddites began to meet in secret (Cherry 2019). Luddite Councils are named in honor of General Ned Ludd, a stylized and embellished figure who worked as a child laborer in nineteenth century textile mills. In "The Future Encyclopedia of Luddism", author Miriam A. Cherry writes that at the age of 15, after refusing to work on a dangerous textile machine and being docked several day's wages, Ludd took a hammer to the loom and smashed it. He fled to join the military, returning a short time later hoping to organize the factory workers. He formed a grass-roots coalition of laborers that eventually was associated with its leader, the Luddite Movement. It was open to all, with roughly 40 percent of the Luddite membership made up of women and children and growing to over 100,000 by 1811. They held a deep-seated opposition to the use of machinery in the textile industry, believing that an increasing reliance on machines was a threat to product quality and ultimately diminished the skills of workers.

Figure 9 Nedd Ludd, hand-colored etching, 1812, Granger Collection, New York

Luddites engaged in a variety of protest tactics including the destruction of machinery. Their tactics often involved breaking into factories and smashing or damaging the machinery in the hopes of both disrupting production and sending a message to factory owners. The tactic was met with fierce resistance from factory owners and government leaders, including an act passed by the House of

Commons “punishing with death those who destroyed the machinery used in making cloth or hosiery of woolen materials” (Felkin 1887, 228).

The Luddites leveraged these acts of resistance, intimidation, and violence by targeting their threats and violence at factory owners, supervisors, and those individuals perceived as supporting the industrialization of the lace industry (Palliser 1869). In his 1887 book *A History of the Machine-wrought Hosiery and Lace Manufactures* (1887), author William Felkin states that the times “became troublesome and dangerous” (Felkin 1887, 227). In June 1816 there were two attacks which destroyed nineteen lace machines in one factory and fifty-five frames in another. Two men were tried for the first offence, a crime that carried a potential sentence of death. They avoided conviction by successfully using an alibi from a fellow Luddite whose membership was unknown to authorities (ibid. 1887). The members accused in the second factory ransacking also won an acquittal from their trial, but one member was convicted of shooting at a guard and was executed. In 1811, Luddite rioters disguised in women’s clothing smashed power looms and burned down factories. Many of the participants were later apprehended and eventually hanged (Waller 2016).

Luddite members shared loyalty and solidarity with their brothers and sisters in the movement, with the organization’s leadership and its political and social aims held as closely guarded secrets (Cherry 2019). The public response to the Luddite movement was mixed, with some workers and community members supporting the efforts, while others viewed their actions as a threat to social order and the economic progress of the country (E. Jackson 1900). Thomas Gardiner, a lace mill owner, argued that support of the Luddites fight against mechanization was misplaced, stating “I am sorry to find well educated persons join them in saying they are injurious to their interests. Genius is not to be stopped in this savage manner” (Felkin 1887, 237).

The craftspeople were angry about a future where machines would take away their livelihood, feeling powerless to effectively adapt and maintain their quality of life. In time, it became clear that the Luddites’ railing against technological progress was futile, this struggle between labor and invention would not land in favor of the workers. Thomas Carlyle, Charles Dickens, and Karl Marx saw the mechanization of manufacturing as dehumanizing and spiritually impoverishing (Waller 2016). Felkin describes the rise and fall of the Luddite movement as a “wretched heap of wrongdoing”, identifying the hunger and misery that resulted from industrialization as the catalyst for their anger. An anger from which they never fully emerged, even forty years later (Felkin 1887, 239).

With the rise in steam powered factories, corporate profits soared. It took over fifty years for the workers who fueled the revolution to see their real wages rise (F. Engels 1892). This decades-long


Figure 10 Friedrich Engels (1820-1895)
Historisches Zentrum, Wuppertal

time-gap, marked initially by innovation and job loss and capped by job creation and recovery, was described by a twenty-four-year-old German author Friedrich Engels. First presented in *The Condition of the Working Class in England in 1844* (1892), his theory detailed the significant span of time between when a technology is introduced and when new jobs are created, jobs that offset those jobs lost in previous decades, with a corresponding rise in wages. Economists now refer to this offset of time as “Engels’ Pause”, a term coined by economic historian Robert C. Allen to describe the period from 1790 to 1840. Allen argues that the surge in inequality identified by Engels is an intrinsic part of an economy’s growth process. He points out that this two-stage evolution of a balanced economy involves a beginning phase where the real wages of workers stagnate while the output per worker expands (Allen 2009). With the introduction of steam power, the profit rate of the corporate employers doubled with a resulting expansion of national income, all at the expense of labor and land. It was not until after the middle of the 19th century that wages began to grow in line with worker productivity, and the profit rate stabilized (F. Engels 1892). Allen argues that this surge in inequality was an intrinsic phase of the growth process. The adoption of a technology like automated lace production increased demand for the capital used to implement the innovation, ultimately raising the profit rate and capital’s share. The rise in profits, in turn, sustained the steam-power revolution by financing the necessary accumulation of capital (Allen 2009).

It has been argued that the cycle of wages stagnating as corporate profits soar is a defining metric of the social and economic transformation that is underway in contemporary economies (Acemoglu and Restrepo 2020). Author Carl Benedikt Frey contends that advanced Western economies have entered a new Engels's Pause, drawing comparisons between the introduction of steam power and today’s rise of Artificial Intelligence (Frey 2019). Economists at the Demos Helsinki think tank have also proposed that the widening gulf of inequality in the global distribution of wealth, ever-advancing technology, and an increasingly dynamic labor force, resemble the trends observed in Engels's Pause (Demos Helsinki 2021). Bank of England Governor Mark Carney is concerned that the rapidly increasing technological requirements of blue- and white-collar jobs will result in poor wage growth and worker redundancy (Carney 2018). Economists have noted that the displacement of workers and the capacity of current systems to maintain job development appear to be weakening (Hautamäki, et al. 2017). As we will see, the pattern of innovation resulting in lost jobs and depressed wages, with a recovery measured in decades, is not unique to the first industrial revolution. The cottage workers of handmade lace are but one group who have felt the direct consequences of technological advancement and the rush to automation.

2.5 The Age of Science and Mass Production and The Digital Revolution

A commonly observed result of the technological revolutions of the last 300 years is the creation of “Lost Generations”. Author Kevin Roose writes that these generations are the “millions of people whose lives are derailed by forces outside their control, and who never make it to the promised land, or even live long enough to know what the promised land looks like” (Roose 2022, 19). These people were the grist caught in an economic system that required a “systemic imperative to reduce production costs in relation to prices” for goods and services (Srnicek 2017). Nick Srnicek writes that this drive to undercut markets results in a system that continuously optimizes process and productivity through technological innovations like the “Iron Man” in a quest for profits and capital accumulation. The role of the worker in this paradigm is often viewed as a peripheral concern. The Steam Era was an age of machines. This was a system that depressed the demand for labor because of the reduced number of workers required for manufacturing, resulting in oversupply. The consequence of this oversupply in labor was that workers were easily replaced. It is likely the challenges of oversupply compounded familiar problems brought on with every technological leap. It is likely that these members of the Steam Era’s Lost Generation struggled to adapt because available jobs required they relocate near a factory, they were unskilled for the new jobs created, or they may have been reluctant to adopt the culture and requirements of an unfamiliar work environment, abandoning the skills they had previously earned.

The end of the steam era led to a period of American history marked by staggering corruption, bloody labor clashes, bitter racial injustice, and soaring income inequality (Roose 2022). The Gilded Age delineates the start of the next industrial revolution, The Age of Science and Mass Production. Like the other two, it illustrates technology’s impact on society and the dislocation of entire generations. This age brought tremendous innovation including the invention of gasoline engines, airplanes, and electric power. Innovation was driven by scientific discoveries that offered a promise that you could go faster and accomplish more. In this era of discovery, the invention of the assembly line is one illustration of how scientific principles were incorporated into manufacturing. Henry Ford was an American industrialist, magnate, and founder of the Ford Motor Company who was passionate about creating an efficient assembly process for his Model T cars, breaking each step of fabrication into individual tasks assigned to a worker (Eschner 2016). By the 1920's the company was mass producing millions of Model T cars each year, a revolutionary automobile powered by a gasoline engine that was constructed on the assembly line, all products of the era of mass production (Schwab 2017).

This push towards automation resulted in a mass migration as people followed the opportunities for work in America's cities. The early 1900s saw workers uprooting from their rural homes and moving to factory jobs, resulting in 40% of the US population living in the cities, compared to just 6% in 1800 (Schwab 2017). To the residents of urban areas, it must have appeared that they were witnessing the dawn of the modern mechanized world. The cities were the first to implement sweeping technological inventions like electric lighting, radio, and telephones, transforming the way people lived and communicated.

Yet even with the wonders of innovation, concerns were raised whether job creation would keep up with worker supply. In 1928, New York Times reporter Evans Clark wrote that through increased production and the expansion of ancillary industries automation had created enough jobs to absorb displaced workers. He noted that while this was accomplished without "disrupting society's peace of mind. There is no assurance that this happy balance can always be preserved" (Clark 1928). His article focused on expert predictions that the "new invention" of electric-powered factory machinery would soon make manual labor obsolete (Roose 2022). He feared a consequence of this automation would be a reduced labor force, noting that if the "factories of America turn out their full quota of human supplies and use fewer human hands in the process the machines must take the blame" (Clark 1928). Clark described what he saw as the paradox of prosperity and unemployment, questioning if the growing pre-Depression era unemployment roles were a result of automation. "It begins to look as if the onward march of machines into every corner of our industrial life has driven men out of the factory and into the ranks of the unemployed" (ibid. 1928, 1).

The impact of mass production in England is illustrated by examining how companies were valued. British engineering had become capital intensive rather than labor intensive. Likely stemming from the successful business models developed during the steam-powered era, businesses were far more dependent on owning expensive equipment, relying less on a highly skilled workforce. David Waller argues that this continuation of the marginalization of labor deepened the roots of the UK's notoriously poor industrial relations (Waller 2016).

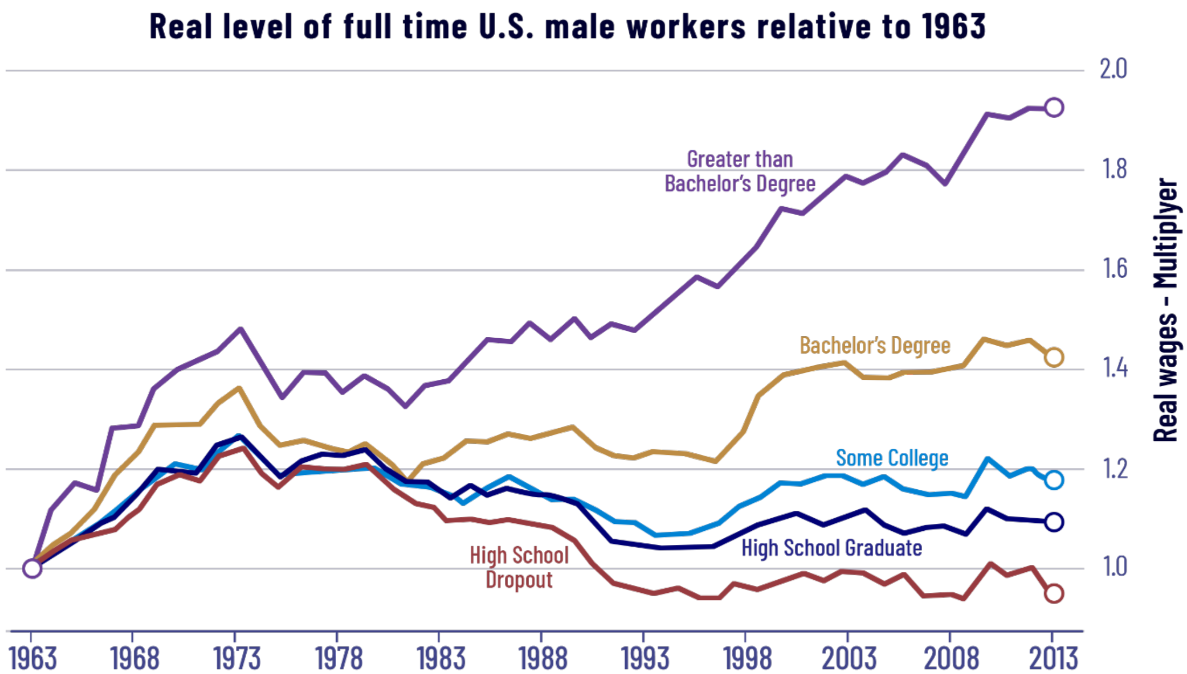


Figure 11 Change in real wages over 40 years.
Adapted from David H. Autor, (2014) and Mark Carney, Bank of England (2018)

This pattern of worker displacement was repeated during the Third Industrial Revolution: The Digital Revolution. Beginning in the 1950s, technical innovation brought semiconductors, mainframe computing, personal computing, and the Internet into both the workplace and homes. This embrace of all things silicon created jobs that required specific technical knowledge. Assembly line workers were replaced by automation, accountants abandoned their paper legers in favor of computer software, postal carriers delivered junk mail because email replaced letters, musicians were replaced by sequencers, and the list continues. From 1987 to 2017, worker displacement dramatically outpaced reinstatement, with the new jobs that were created classified as high-skill jobs requiring qualifications that many workers couldn't access (Autor 2014) (Acemoglu and Restrepo 2020). In the past, when a displaced worker's position was eliminated, they could take solace in the belief that a new job would soon be created for them. Unfortunately, many soon realized that the jobs being destroyed by automation would never return. Where before a worker might be operating a lathe or cold press or spot welder, many of these new jobs required technical skills for operating the automated robot that had replaced them. Like the hand-made lace makers of the 18th century, the skills they carried were no longer valued in the marketplace.

To secure and keep the type of job that offered the prospects of a long-term living wage, a worker in the Digital Revolution likely needed a college education (*see figure 11*) (Autor 2014). Beginning in the 1980’s, opportunities were diminishing for someone seeking to enter the manual labor or unskilled worker labor market, and for those displaced workers looking to replace jobs lost to automation. More importantly, the real wages paid for the manual labor jobs that were available were shrinking while the economy was expanding. A laborer had less money to purchase the goods and services of daily life, goods that were becoming more expensive each year. The jobs created by the shift to digital largely tended to be white-collar jobs requiring a college education (Carney 2018). In the digital revolution the paths for opportunity and remuneration diverged sharply between those with only a high school diploma and those who were college educated. The new jobs created required technical fluency with digital tools, and the ability to acquire new skills as the digital economy evolved.

2.6 Unexpected advances

Each of the industrial revolutions explored in this chapter resulted in improved labor efficiency. As a result, workforce demand fell. While this displacement had significant near-term microeconomic impact on the people who saw their jobs and careers disappear, the long-term macroeconomic result was a boost in productivity and the creation of new demands for labor elsewhere (Roose 2022). Whilst Evans Clark expressed his concerns about increasing unemployment in the 1920’s, America never suffered from the mass unemployment he feared, at least not directly attributed to the technological advancements in Era of Mass Production. Also noteworthy about the three transformative eras is that many of the jobs created by the upheavals were inconceivable prior to the innovation.

The 20th century digital revolution reveals a plethora of jobs created in response to technical advances, from app developers to social media managers, podcast producers to IT Managers, cellphone technicians to drone cinematographers. These heretofore unheard-of careers naturally evolved during the disruptive transition from analog to digital, most in response to problems and opportunities presented by the invention of new technologies. The act of invention is often a process of combinatory creativity, with solutions taken from one problem inspiring a solution for another issue in an entirely different realm. Most of these combinatory iterations are small solutions that solve an immediate problem. Others are significant applications that advance our thinking. A rare few are what might initially appear innocuous but are solutions that radically alter the path of history. One example is the Jacquard card.

The Jacquard card system was used to direct a loom as it wove incredibly complex and detailed patterns. Developed in 1805, the Jacquard loom could produce textiles in a fraction of the time that it would take a master weaver to create the same cloth manually. The loom used a binary system with a sequence of holes punched into a paper card to record a particular woven pattern. The cards stored instructions that could be read by the loom to replicate complex woven fabric, not unlike how a player piano is directed to play specific notes at a specific time by an advancing paper roll (Elliott 2017).

It was the application of the Jacquard weaving system to John Heathcoat's circular bobbin net frame in 1837 that signaled the final blow to the handmade lace makers. With programmable automation, machinery lace like the complex black silk net "dentelle de Cambrai", an imitation of Chantilly lace, could be produced consistently with a quality that matched the work of human lace makers (Palliser 1869). The Jacquard card system afforded all forms of textile production to operate at scale, revolutionizing the manufacture of patterned textiles by enabling the mills to produce them at a fraction of the cost (Elliott 2017).

At the same time English mathematician Charles Babbage was developing a mechanical calculator called the Difference Engine, an automatic mechanical calculator designed to tabulate polynomial functions. Some of the most foundational mathematical functions used in engineering, science, and navigation, are built from logarithmic or trigonometric calculations which can ultimately be represented as polynomials (Kafai and Margolis 2016). Babbage collaborated with Ada Lovelace, a mathematician and author, who recognized that the binary system used with the Jacquard loom could be applied to their analytical engine. To operate the difference machine a user needed to provide instruction for which mechanical switches would be turned on an off, instructions that could be programmed into paper cards like the Jacquard system. Lovelace understood the direct connection between the loom and their engine, writing "the Analytical Engine weaves algebraic patterns, just as the Jacquard loom weaves flowers and leaves" (Elliott 2017).

Figure 12 Ada Lovelace, Wikimedia Commons/Public Domain

She was a prolific analytic writer, creating an elaborate series of "notes" from her work translating an article responding to their Analytical Engine proposal by Italian military engineer Luigi Menabrea. Her notes included an algorithm designed to be executed by their machine for calculations that some contemporary computer scientists have described as the first computer program, with Lovelace named the first programmer (Fuegi and Francis 2003). Lovelace wrote about her vision of computers being more than a calculator, describing her mindset of examining how individuals and society relate to technology as "poetical science" (Lovelace 1842).

In one rather prescient view of contemporary discussions, Lovelace dismissed the notion of a computer having artificial intelligence, writing that "the Analytical Engine has no pretensions whatever to *originate* anything. It can do *whatever we know how to order it* to perform. It can follow analysis; but it has no power of anticipating any analytical relations or truths" (ibid. 1842). Her dismissals of AI's potential have been extensively reviewed by computer scientists, including Alan Turing, who observed that Lovelace appeared to argue that “a machine can ‘never do anything really new’” (Turing 1950). He did offer the qualification that for Lovelace, imagining a computer able to “think for itself” would require designing a computer that demonstrated a conditioned reflex to serve as the basis for learning, and it did not appear the machines “constructed or projected at the time had this property” (ibid. 1950).

The collision of lace, punch cards, Ada Lovelace, Alan Touring, and Artificial Intelligence is a significant story of iteration and the interconnectedness of human intelligence. It is noteworthy to trace the evolutionary progression from 18th century handmade lace, to 1804's introduction of the binary Jacquard card system, to lace automation in 1837, to the Difference Engine in 1843, to 1950's mainframe computer's use of punch cards, to today's continued reliance on the same system of ones and zeros that started in a textile mill in France. A system that very well could displace hundreds of thousands of workers worldwide via Artificial Intelligence.

2.7 “Boomer remover”

The specters of job loss, displacement and depressed wages loom menacingly over any Artificial Intelligence discussion. There are well founded fears that labor from the full spectrum of industries could suffer the same fate as the handmade lace makers, inexorably squeezed out of a livelihood by a rapidly evolving technology. Many hope that we will strike a sort of occupational detente where craftspeople and AI work side by side. Human-generated and machine-generated creativity did manage to coexist in the early stages of automated lace production. Elaine Freedgood notes that, “handicraft and machine production of lace once resided unproblematically in the same word: ‘manufacture’” (Freedgood 2003, 629). In the 16th and 17th centuries, the term 'manufacturing' referred to the manually intensive production of goods, either through physical labor or utilizing mechanical power. As those who “manufactured” lace by hand eventually learned, as the technology improved and the cost of production went down, their manufacture lost value. This arc of the introduction of a new technology, a window where old and new coexist, and the eventual displacement of old with new presents the question, is it likely that we will see a similar outcome for workers who are finding themselves in the crosshairs of AI? Additionally, if this arc of technology is repeated, who is likely to feel the first impact?

Initial implementations of Artificial General Intelligence have been primarily aimed at solving problems where failure is expensive. For example, the development of self-driving trucks, cars, and forklifts require a level of safety where accidents are considered intolerable. With recent use cases like graphic design and music composition, the stakes are no longer life-or-death. A creative and expressive tool like DALL·E 2 that generates some great pictures, but also some clunkers, is considered tolerable. This shifting focus to acceptable levels of failure is opening a new world of applications (Mollick 2022). AI is adept at modeling jobs like production planning, a process where factories calculate which machines in their plant should be producing a particular item at a particular time. Corporations spend billions of dollars each year paying human analysts to pour over resources like demographics and sales data, projecting customer demand and scaling production accordingly. These projections are a critical component to the success of a company. The consequence of introducing AI tools that perform millions of virtual simulations in only minutes to execute the same analysis is likely to make human analysts expendable. This would afford factories the ability to replace entire teams of human production planners, along with most of their outdated software. Author Kevin Roose quoted one AI developer describing their production planning product as “The Boomer Remover”, celebrating that their product would allow companies to eliminate the “old, overpaid middle-managers who aren’t really necessary anymore” (Roose 2022, xvi). Journalist Phil Edwards also described AI’s potential impact, noting, “These machines are the stuff that would put people out of work” (P. Edwards 2022, 02:55).

The first 100 years of automation focused on repetitive manual tasks with mechanical tools. The displaced workers were concentrated in blue-collar manufacturing jobs, while white-collar knowledge workers largely considered themselves safe. What has changed with the introduction of AI is White-collar workers are more likely to be automated out of a job than blue-collar workers (Webb 2019). Many of the most promising and talked about applications of AI are in the professions of accounting, law, finance, and medicine (Roose 2022). In a 2019 report, job descriptions and AI-related patents were analyzed by the Brookings Institution to investigate potential intersections of key words. Using data from Stanford researcher Michael Webb, phrases such as "predict quality" or "generate recommendation" featured prominently. Of the 769 job categories included, Webb and the Brookings researchers uncovered that 740 of them were vulnerable to automation in the near-term. Notably, workers with bachelor’s or graduate degrees were at a much greater risk to feel the effects than those with only a high school education--with the college educated nearly four times as likely to feel AI’s impact in the short term. The study found that some of the most automation-prone jobs were in highly paid occupations in major metropolitan areas like San Jose, Seattle, and Salt Lake City (Webb 2019).

Start-ups are raising millions of dollars in venture capital to apply generative AI to areas like gaming, programming, and advertising (Roose 2022). Three of the current strengths of machine learning are pattern recognition, data analysis, and iterating variations of a specific item, tasks that are an integral part of fashion design. "Fast-fashion" chains and e-commerce brands like Glitch are already implementing fashion AI for generating custom designs for customers (*figure 13*) (Dozier 2019). A team at Amazon developed an algorithm that analyzes images of garments in a particular style, learning how to generate new variations (Ong 2017). The AI-based music-writing tool Jukedeck allows users to generate new compositions on the fly. Journalist Clive Thompson reported that the impact of the tool could result in deep cuts into the work of studio musicians who create soundtracks and stock music libraries (C. Thompson 2019). Adobe has integrated their Sensi AI into the motion graphics application AfterEffects, dramatically reducing the tedium of frame-by-frame outlining when rotoscoping a moving image. They have also included Sensi tools like "reframe" and "scene detection" in their video editing platform, PremierePro, for automatically centering a video's subject matter in the frame and making edits for each shot change (Adobe 2023). ChatGPT is being used to replace specialized tasks in creative crafts because it excels at technical assignments like writing AfterEffects expressions. This type of instruction code is used to control an animation's graphic elements. Because of their complexity, writing AfterEffects scripts by hand often takes many minutes, or even hours, to write, modify and perfect.

Figure 13 Glitch Fashion, June 2019

Some creatives have expressed appreciation for the power of these tools. The goal of most of these applications is to eliminate the tedium of repetitive tasks that break an artist's creative flow. Designer and educator Yannick Theunissen described his experience of using ChatGPT to write AfterEffects scripts, marveling that "your role in creating the graphic is now merely supervisory" (Theunissen 2022). OpenAI views DALL·E 2 as a sketching tool to augment and extend human creativity, not replace it (Paetzhold 2022).

Others fear that the introduction of AI into the creative workspace will be devastating to designers, believing that no human illustrator can work fast enough or cheap enough to compete with these robot replacements (Crabapple 2022). The applications generate images and music exponentially cheaper and faster than their human counterparts. Amazon's development of the fashion algorithm generated concern that if Amazon spots a trend and quickly designs on-demand clothes, their reach could create a clothing steamroller that could flatten other clothing retailers (Ong 2017).

There appears to be no such thing as an inherently robot-proof job. Still, creatives are strategizing how they can possibly compete against a system that works at unfathomable speed, generating multiple variations of an idea, all at pennies on the dollar. Most are searching for what elements of their craft differentiate their work from AI, elements that display their "humanness", believing that AI generated works "remove the humanity from the art process" (Andersen 2023, 28:31). They are hoping that their ability to capture nuance and feeling, to define a clear point of view, to express what they are feeling through their art, will matter. Hollie Mengert, a Disney illustrator, points with pride at her ability to create "authentic expressions, appealing design, and relatable characters, and I feel like that is something that I see AI struggle with most of all" (Baio 2022). Generative AI is very good at learning patterns, mimicking the brush textures, colors, and shapes within a particular genre or style. Artists believe merely replicating those brushstrokes and colors results in artwork that is superficial, it lacks a genuine point of view formed during creation. For many, it matters how art is produced because they believe the process of creation adds significant value for both the artist and those who experience it (Eisikovits and Stubbs 2023). Part of what gives creative output its power is the process of witnessing someone's natural gifts on display. People enjoy and celebrate human creativity because it represents the epitome of human achievement, the fusion of talent, human gifts, and hard work. Many see art as a captivating expression of humanity and its inherent originality. It reflects the cadence of human labor, thought, and dedication that creates compelling works inspired by any number of sources, be it life experience or imagination, rather than the homogenized voice of AI.

An artist's chosen craft will not necessarily determine if they are replaced by a machine. Instead, what may matter most is how they approach the task. Evans Clark described the anxieties of those who worried about their employment in 1920, poetically capturing the angst and ambiguity of their uncertain future.

> The full story of men displaced by machines has yet to be written. For each individual, of course, it is a unique saga of hopes and fears, disappointments, and, sometimes, fulfillment. But the main outlines of the plot in each case are much the same (Clark 1928).

Like tailors of hand-made suits and cobblers of hand-made shoes, the work produced by the creatives left standing could serve as status symbols for marketing teams and studios. Others see potential to use the new AI tools to enhance their creativity, to find a new voice as a storyteller, and to jump-start the creative process. They point to the notion that AI image generators have the capacity to create images that are beyond their imagination, pushing their boundaries and capabilities, "providing the tools I need to do my best work" (Theunissen 2022, 06:03).

As AI's use expands in the creative crafts it is uncertain who will stay and who will go, who will command payment for their work and who will need to find an alternative career, who will have the opportunity to craft stories and who will be relegated to watching them. Some artists take the position that only a tiny elite will remain in business.

2.8 AI Hybrid

When the daguerreotype was introduced in the 1800's, portrait painters believed the camera would destroy their livelihoods. Others saw the new technology as a path of expression, embracing the new medium's possibilities. As the camera was established as a viable way to capture a person's likeness, painted portraiture's popularity ebbed, but the medium of painting persisted. Artists pivoted, finding new aesthetic problems to solve (Sutton 2022). Image capture changed again in the 1990's with photography's digital transformation, marking a shift from film to files. Computer applications like Photoshop afforded photographers the means to create images in new ways, producing photos they could not replicate in the darkroom. Today's shift from digital processing to AI image generators differs significantly from the introduction of digital photography 20 years ago. The results achieved with a tool like Photoshop rarely surprise the photographer, where the output from a tool like Stable Diffusion are a consistent wonder.

An ever-growing number of fine artists are incorporating Artificial Intelligence into their creation process. These artists are embracing AI technology to support, enhance, simulate, or replicate creative processes, thereby indicating that new artistic forms are emerging with AI used as both a tool and the subject of artistic works (A. Webb 2022). This exploration of aesthetic outcomes is happening as artists work in concert with the technology by generating and modifying parameters of a machine learning model, supporting their creativity (A. D. Thompson 2022). AI tools are enhancing artistic expression and creativity, serving as a catalyst for new ideas. They are pushing established artistic boundaries, with a likely outcome of not just an expansion of fine art, but the follow-on effect of commercial art applying these aesthetic discoveries to marketing and advertising. Many of these aesthetic innovations will inevitably be appropriated by marketing teams, percolating into popular culture. Nir Eisikovits and Alec Stubbs write that the link between human ideation and execution does not need to wither with the introduction of AI. If the applications are viewed as mechanisms for creative imagining, what OpenAI refers to as "extending creativity", they can be used to generate provocative stimuli, prompting artists to conceive of innovative work through imaginative thinking (Eisikovits and Stubbs 2023). Much of the concern expressed in popular media about AI's impact on the act of creation is framed by a perceived loss of meaning (Brooks 2023). There is an inherent perception that AI images feel shallow, lacking a depth of human emotion or point of view, rendering the artistic process meaningless. Eisikovits and Stubbs believe that for a creative outcome that incorporates AI to "count as art", artists will need to perform the bulk of the work.

Nearly 100 years before the introduction of DALL·E, William Ralph Inge wrote in *Assessments and Anticipation*, "What is originality? Undetected plagiarism" (Inge 1929). It has been argued that human creative products start from the imitation of others because humans are imitative creatures (T. E. Jackson 2017) (Turkle 2005). Nothing comes from nowhere; creative pieces are a combination of what has come before. Therefore, no work is "original" in its purest form. Author Austin Kleon writes in his book *Steal Like an Artist,* that "every new idea is just a mashup or a remix of one or more previous ideas" (Kleon 2012, 9). Originality is an abstract concept and there is rarely work that is entirely new. The professional artist's understanding of this reality is what separates them from amateurs. If an artist steps away from the assumption that they must be completely original, abandoning a belief that they must create something out of nothing, then they are free to respect the influences of their creativity.

Does AI artwork fall into Kleon's definition of art as a mashup or remix of previous ideas? Mingyong Cheng asserts that since AI-generated images are a combination of previous work that presents new ideas in surprising and unpredictable ways, then they exhibit the original thinking of human artists (Cheng 2022). You can apply Keith Sawyer's assertion that creativity is a combination of existing ideas to AI artwork (Sawyer 2012), and also Deniz Kurt's belief that thoughts and concepts are a composition of thoughts and concepts that already exist (Kurt 2018) Margaret Boden writes that ideation and creativity simply involves, "making unfamiliar combinations of familiar ideas" (Boden 2004, 3), which can be applied to the actions of both human and AI artists. Therefore, whenever the human artist uses AI for creative purposes, pushing beyond the mechanical augmentation of algorithms, both the act and the tool play significant roles in influencing the artistic process (Arriagada and Arriagada-Bruneau 2022).

This act of combining previous ideas to create something new, “making unfamiliar combinations of familiar ideas”, was described by Boden as combinatoric creativity (Boden 2004). The theory describes our ability to spot a problem in one area of our life, and apply knowledge gained from an entirely different area to fix it. Maria Popova illustrates the concept by pointing out that we might apply a piece of advice given to us in middle school to an unexpected life situation decades later (Popova 2011). On the surface this theory of combinatoric creativity appears to also describe AI’s ability to generate output that is new and unique. It is important to note that a significant element within the application of this idea is our ability to randomly access information across the entire breadth of our knowledge. Each of us hold a vast repository of random, disparate information in our memory, available to be combined and synthesized at a moment’s notice. What appears to us to be a “bolt out of the blue” is a clutch of seemingly serendipitous associations accessed through random synaptic connections that facilitate the idea. By melding different concepts to form our ideas, we create new combinations unknowingly or knowingly (Kurt 2018).

One challenge with current AI models is that while they can combine what has been learned from a specific data set to fabricate an answer, they are unable to make random connections across multiple disciplines to produce something that is truly novel. They suffer from a form of linear thinking that restricts their capacity to produce the type of results that synthesize learning across multiple disciplines and skills. Boden argues that while combinational creativity is the most straightforward AI-driven display of creativity, it is also the most problematic. To be considered a successful creative artifact, a valuable creation, it must possess a degree of relevance that can be both understood by, and meaningful to, a human audience (Boden 2004). The aptitude for exploring unfamiliar combinations of familiar ideas as been *the* defining asset for human creators in the idea economy. Should combinational creativity remain a uniquely human skill, if AI is unable to autonomously facilitate random combinations across multiple disciplines, then creatives and artists will have the means to differentiate themselves from machine-generated output.

Creating work without leveraging any AI tools is one path of differentiation. Unfortunately, that choice is perilous because the creative artist will be competing in the marketplace with machines that are both prolific and cheap. The most likely alternative is to leverage AI as only one of many tools in a creative arsenal. Still, that choice will require the creative to perform a balancing act. For their AI-assisted artwork to present a point of view, a perspective that will resonate with their intended audience, creatives will need to go beyond the act of simply writing prompts into an image generator. Instead, they will need to develop a personal human-AI hybrid workflow that calculates cost effectiveness while preserving artistic vision. The perceived benefit is that this model has the potential to improve creative efficiency, minimize tedium, expand their imagination, and keep them in business.

Ran Segall notes that as a graphic designer he often receives poor quality images from clients for use in marketing. Rather than use them as is, he leverages the photos as an AI model's training set, ultimately producing high quality product shots. "We don't have to go do crazy photo shoots or hire artists. We can just use our imagination" (Segall 2022, 4:31). Frank Bruni writes that as a journalist he often asks for help and ideas from friends, colleagues, and editors. Incorporating AI into his workflow makes sense to him because, "My field of vision is only so wide, my brain only so big. I'd be a fool not to supplement" (Bruni 2022). Thomas Cargill of Satori Graphics shared that when someone works for themselves they quickly realize the importance of time management, noting that using AI in his process is critical because, "as a designer I am always trying to save time, and to look for ways to save even more time" (Cargill 2023, 3:42). Kevin Roose interviewed Sarah Drummond, a service designer in London, for a New York Times article about how AI-generated art is transforming creative work. She notes that she would not use AI generated images for a client's final deliverable. Instead, she would hire an artist to create media that fully realizes their intent. She incorporates AI as a sketch tool during creative ideation for "the throwaway work that you do when you're any kind of designer, whether it's visual, architectural, or urban planner" (Roose 2022).

The idea of using AI generated media as a starting point for an imaginative journey is a common refrain heard from creatives. Interior designers are using AI to form initial ideas, video influencers use it to brainstorm, and filmmakers are using it to structure stories (Roose 2022) (Brownlee 2022) (Theunissen 2022). Yet each of these instances imply that the personal touch is a critical element of the process. These creatives are using the tools for ideation while retaining the empathetic touch of a human. Human perspectives remain an advantage in the marketplace. There is a compelling need for human intervention with the algorithm to generate media that resonates (Dozier 2019). An illustrator will work with a client as they produce concepts, exploring human experience to communicate an emotion or opinion that will resonate with a viewer (Strickland 2022). Many creatives believe it is important to afford the technology room to improve, while always protecting the spectrum of creative voices. The fear is the mass production of generative output will result in a homogenization of creativity. When 19th century automated lace production swamped the market for handmade lace the Victorian middle class rued the loss of the "human" touch. Despite their enthusiastic participation in commodity culture, the Victorians felt the mass-produced product lacked comparable monetary, social, or aesthetic value. They celebrated the idiosyncrasies previously found in handmade work by buying published catalogs, like Emily Jackson's 300 page *A History of Handmade Lace* (1900), that recorded lace in all its forms (Freedgood 2003). They felt those idiosyncrasies had been destroyed by mass production.

AI automation will reach far beyond the jobs found in creative trades. The jobs that are likely to surface in industries like Information Technology, Planning, and Finance, are also likely to appear in the creative industries. In 2018, the consulting firm Accenture surveyed thousands of large corporations to see what types of jobs were being created as an outgrowth of AI implementation. They grouped the jobs into three broad categories: trainers, explainers, and sustainers. They defined the role of trainers as guiding AI models and refining learning algorithms. Explainers act as intermediaries between the AI models and humans, explaining the decisions and outcomes. Sustainers fill the role of integrating the tools into a company's computer infrastructure (Wilson, Daugherty and Morini-Bianzino 2017). A different list of jobs, created by the consulting firm Cognizant, is far more granular, listing dozens of new jobs. They range in scope from AI business development manager to Master of Edge Computing to Walker/Talker. The latter's responsibilities are to guide both the unemployed and underemployed to find new types of work after being displaced by AI (Cognizant 2017). Both new and existing jobs are likely to fall into the human-AI hybrid model, where an employee oversees most of the work and uses AI as their "helper". These are the type of jobs often described as "human-centered automation" by those who see the prospects of AI in the workplace as exciting, believing that the technology will support the worker rather than replace them.

The first high-demand AI position is already appearing on job search sites, Prompt Designer *(see fig. 14)*. With the release of large language and image generation models, organizations have come to realize the need for employees who can act as intermediaries between human and machine. A Prompt Designer's job is to coax the desired result from an unpredictable algorithm. When GPT-3 was released in June of 2020, it immediately became clear that large language models require a specific approach when writing prompts. This communication is text driven, and the critical components are both semantic context and word choice. A determining factor in the generated media's quality is the nuance of the initial prompt's language. Graphic designer Merzmensch noted that when entering a

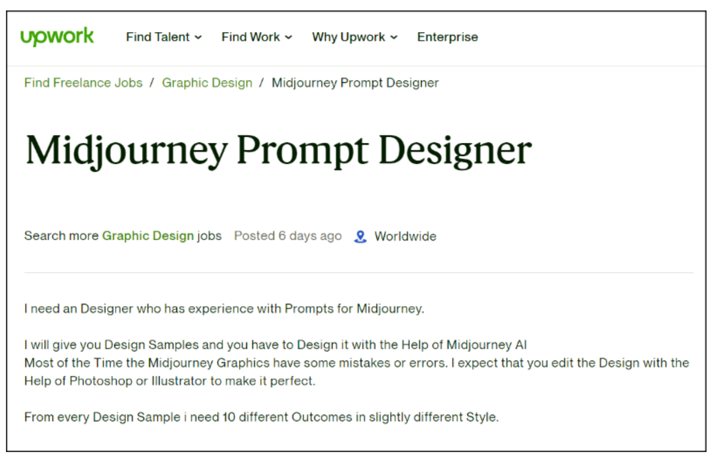


Figure 14 Midjourney Prompt Designer ad on Upwork job search site, February 10, 2023

prompt, one should “neither force it nor request — but inspire, nudge the Transformer-driven model” (Merzmensch 2022). The better the description – the more accurate, the more detailed, the more it uses specific language – the better the outcome (Eisikovits and Stubbs 2023). Successful interaction between human and machine requires a familiarity with the generative psychology of the algorithm. This algorithmic fluency is an amalgam of technology, humanities, and creativity, and because of its ambiguity corporate managers occasionally perceive the practice as a dark art. To be fluent a user needs to leverage the specific modifiers and prompt elements that will determine the quality of the resulting image or text. Once a result is generated by the AI model it is likely a user will need to refine their prompt for a new round, iterating and repeating the process. When the results are ‘close enough’ to meet expectations, the generated media will often need a final human touch, such as text editing, combining multiple results, or cleaning up the artwork.

On the surface, Generative AI models appear to enable anyone with access to realize their imagination. These applications produce media exponentially cheaper and faster than their human counterparts. Viewers often respond to these works in wonderment, expressing something akin to, “AI art looks like an artist made it” (Crabapple 2022). Recent studies have shown that when a respondent does not know an image’s creator they are unable to determine if it is generated by a human or AI (Elgammal 2018) (Cheng 2022). When a viewer is unaware a priori of the source’s origin, AI generated work has ranked higher in what researchers have called “likeness”, or how much they like a piece of work (Arriagada and Arriagada-Bruneau 2022, 94). One example of this preference is a study showing that subjects awarded higher ratings to images created by a Creative Adversarial Network image model than the work produced by human artists, believing all the work was crafted by human hands (Elgammal, Liu, et al. 2017). These generative AI systems consistently produce media in an extraordinary form (Cheng 2022). Because of the public’s positive response to this extraordinary machine-made media, creatives are concerned about the potential devaluation of their work.

As a creative, I believe that the products of my peers fabricate artistic structures that are far more complex than AI-generated media. I have heard concerns that that AI-generated media will replace human-made media with shallow "prima facie" representations, replicating the obvious while lacking the depth of thought or mastery of craft. Much of what is being generated by AI could be classified as ‘decorative art’, as opposed to traditional representations of ‘fine art’ or ‘quality storytelling’. There is a parallel between contemporary machine-made media and 19th century machine-made lace. Neither is likely to represent the emotions and intellect of the creator, nor do they communicate feelings between the maker and observer. Despite this perception that AI-generated

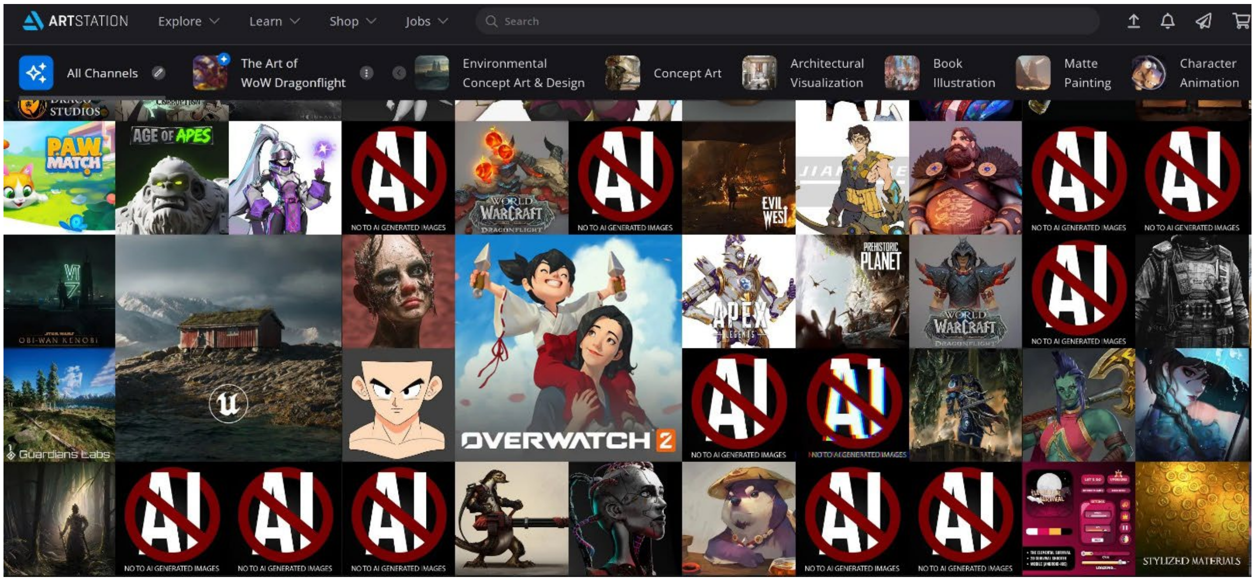


Figure 15 Artstation "AI art free zone" icons from December 2022 (Merzmensch, Artstation: "AI art free zone"… 2022)

media lacks emotional depth, the value of any creation is in the hands of the public's reception (Arriagada and Arriagada-Bruneau 2022). When AI-generated media is 'liked' by observers, meaning media that lacks the assumed depth and complexity of human-created work, it can prompt an existential dread within the creative community. A dread that one's measurement of self-worth and accomplishment may have been a chimera of perceived value.

The response to AI by many creatives in the gaming, film, entertainment, and media industries has been condemnation of AI-generated work (Merzmensch 2022). At the end of 2022, postings on the website ArtStation, a site that acts as a showcase for creatives to share their work, displayed graphics that denounced computer created design (*see figure 15*) (ArtStation 2023). Many of the artists expressed offence with the juxtaposition of AI-generated images and their own work, believing that it degraded and undermined the time and skill invested in their craft (Xiang 2022). Their 'No to AI' graphic campaign reflected a form of human chauvinism, a belief that AI-generated work should not be featured or celebrated within a human-centric community. In an article on The Conversation website, Nir Eisikovits and Alec Stubbs described their belief that "AI art devalues the act of artistic creation for both the artist and the public (Eisikovits and Stubbs 2023). As the artist revolt on the site grew, ArtStation defended the inclusion of AI-generated work, noting that their "content guidelines do not prohibit the use of AI in the process of artwork being posted" (ArtStation 2023). They expressed a concern that acting as a content gatekeeper, checking the provenance of the thousands of images posted on the site, would present an undue burden.

Artists' condemnation of AI-generated work has extended to the writers responsible for most of the television and film scripts produced in the United States. Concern over the potential use of AI to generate scripts and displace professional writers proved to be an intractable sticking point during the 2023 Writers Guild of America (WGA) contract negotiations, leading to a work stoppage. The contract proposed by the union would forbid studios from using AI-generated content as source material and prohibit producers from using a resource like GPT to rewrite human-created work (Pulliam-Moore 2023). In March 2023 the WGA issued a statement on Twitter that presented a clear indication of their position: “Companies can’t use AI to undermine writers’ working standards including compensation, residuals, separated rights and credits,” ending their social media post with, “plagiarism is a feature of the AI process” (Writers Guild of America, East 2023).

A counterpoint to this human-centric view of creativity is that the inclusion of AI-generated media diversifies creative processes and breaks down spheres of elitism, ultimately expanding our definition of art. Gabriela Arriagada-Bruneau and Leonardo Arriagada argue in their article, *AI’s Role in Creative Processes: A Functionalist Approach* (2022), that this sort of democratization is rupturing the artworld’s traditional hierarchical view. Having access to creative tools without facing the barriers of expensive training, costly materials, and the vetting process of distribution via brokers and galleries, is likely to generate alternative creative paradigms. This argument is not without merit. The adoption of video sharing platforms like YouTube not only afforded a cost-free way to distribute content that had been created at little cost, but it also resulted in entirely new formats of storytelling. This was the democratization of a medium that had previously been controlled by the elitist spheres of large studios and television networks, requiring expensive production and distribution. Video sharing sites gave voice to individuals and communities that were previously underrepresented or ignored. The adoption of AI-generated media has the potential to dissolve many of the traditional views of ‘what is art?’ These new forms of AI-generative media, content that combines ideas, styles and views into alternative points of view, has the potential to alter many of the traditions and styles of existing media, just like YouTube transformed video storytelling. The hope expressed by Arriagada and Arriagada-Bruneau is that ultimately this disruption will not displace the creation of human-generated art. That may be difficult to accomplish. A consequence of the rapid expansion of platforms like YouTube has been the reduction and restructuring of traditional media producers, displacing the production facility’s human creators.

Many artists believe that AI is incapable of generating authentic human imagination, creativity or perspective, that it is simply a tool designed to process data and perform specific tasks (Brownlee 2022). They argue that the reason for the illusion of quality is they are fabricating derivative creations through regurgitation, replicating the style of the artists used for training. These tools exist because of the music, photos, artwork, graphics, and drawings scraped from the internet and consolidated into massive data sets like LAION-5B, a training set with 5.85 billion CLIP-filtered image-text pairs (Schuhmann, et al. 2022). These training data sets are filled with work created by humans, reflecting inherent human perspectives. Yet like the Iron Man who replicated the work of 18th century laborers, these AI models work longer, faster and do not take vacations. For many organizations, the cost effectiveness and efficiency of utilizing these automated tools is seen as good business. Reducing the expensive overhead of a creative workforce through staff reductions affords them an avenue for creating content that is likely to be judged by their customers as 'good enough'.

2.9 الكلاب تنبح والقافلة تسير "The Dogs Bark, but the Caravan Moves On"

The evolution of Artificial Intelligence and Machine Learning is accelerating at a nearly unfathomable rate. It is safe to say that the technology has arrived, and we are already feeling its effects. It is doubtful that an individual creator can carve out a path where little of their workflow changes. Like the craftspeople who manufactured handmade lace, they run the risk of displacement. The scale of innovation and capital investment means there is little these individuals can do to alter the global scope of the AI Revolution. Their expressions of anger and frustration, while valid human emotions, are unlikely to have a noticeable impact. An Arab proverb states, "The Dogs Bark, but the Caravan Moves On" (Ghuloum 2015), a somewhat fatalistic description of the creative community's protestations and their likely impact on the rate of adoption.

Many of the companies developing AI are focused on pushing the boundaries of AI. They are focusing on developments like diffusion models for full motion video, autonomous taxi cabs, or creating efficiencies in their financial analysis software in the hopes that it can do the work of 20 accountants (Roose 2022). While many corporate leaders express a desire to embed efficiencies into their applications as a means of affording individuals more productive use of their time, the impact of these efforts will almost certainly lead to the displacement of many, if not millions, of workers. It is an arms race between multinational companies and nimble start-ups to own market share. Some corporate leaders are preaching a hands-off approach when considering how their technology will be used. Emad Mostaque, the founder and chief executive of Stability AI, is freely sharing the technology with millions of users as Open-Source Software, with little consideration for potential outcomes (Roose 2022). Journalist Maggie Harrison describes how this free-for-all deployment of AI is lacking any forethought about the potential harm to both individuals and society, writing that "AI is still an experiment — and collateral damage isn't a prospective threat. It's a given" (Harrison 2022).

History shows that technology has a record of producing unforeseen outcomes. While the Steam Age introduced manufacturing on a previously unimagined scale, hundreds of thousands of workers were feeling the consequence of automation in lost wages, collapsing local economies and ways of life forever altered. Some technologies take much longer to present unforeseen consequences. Jacquard introduced his punch card system in 1804, which had a direct impact on Babbage's difference engine, which led to IBM punch cards in 1928, which led to Mainframe computing in the 1960s and ultimately to today's AI revolution. Innovations may push users in one anticipated direction, resulting in totally unforeseen outcomes a few years later. An example is the introduction of high-quality cameras in mobile phones. These cameras led to an unexpected drop in demand for professional photographers, and ultimately resulted in the near total collapse of the largest film manufacturer in the world, Kodak. This brief list of unpredictable, yet consequential, technological outcomes underscore my belief that no one can confidently predict the future of AI. Innovation can be painful, so the challenge for creatives is how does one adapt?

One path of adaptation may rest on one's humanity. The most valuable ability in the coming AI economy is likely to be refining the skills that differentiate an individual from a machine. One's worth in the marketplace is likely to be measured by unique human abilities that machines are unable to replicate. As we have discussed, these AI models replicate patterns rather than emotions. The creative output is initially appealing to observers, but often lacks emotional and intellectual depth. In practical terms, those who can effectively communicate social desires and project human emotion in their media are likely to prove difficult to automate out of a job. The workers who are most at risk appear to be those who perform tasks that are measured by efficiency and repetition. Successful creatives are likely to weave these AI tools into their portfolio of skills. The most effective creators in the non-AI economy intuitively understand that their success comes from their ability to generate emotional responses from an intended viewer. Defining and presenting that skill will be their most significant market differentiator going forward. Thomas Cargill believes that "there is always going to be a need and a market for human-made design. Brands will want to hire human designers and will pay more for that. But there will be less of those designers" (Cargill 2023, 6:49).

AI models are very adept at operating in static and stable environments, expected prompts, and clearly defined rules. The human creator's advantage over machines is an ability to manage ambiguity, to respond to change, and to invent new rules along the way. Kevin Roose points out that none of these questions concerning the impact of AI are about machines. “They’re all about people” (Roose 2022, 22). AI is the catalyst for the coming upheaval, but a question we need to answer is, how will humans learn to interact, control, and respond to the challenge? Evans Clark understood this inherent tension between progress and peril when he wrote in 1928, “For the God of the machine, ultimate perfection is a world which runs by itself, mechanically, with only here and there the human touch. But such a world would require infinite social readjustment” (Clark 1928).

Developing a thinking machine, an effective model of human cognition that embodies Clark’s ultimate perfection, has proved to be an elusive quest. While AI may represent the next disruptive industrial revolution, its time of development spans the previous two. Clark’s words are proving prophetic, as it appears our adjustment to the consequences of Artificial Intelligence will truly require “infinite social readjustment.”

Chapter 3

AI models and implementation in creative computation

Is the human mind computable? The rush to develop a systematic and effective model of human cognition began in earnest after the Second World War. Computing pioneers like Alan Turing, Douglas Rayner Hartree, Frank Rosenblatt, John McCarthy, Marvin Minsky and Ray Kurzweil have each altered the course of development. Tracing the arc of AI from Turbochamp to the Perceptron, from symbolic models to neural models, from Narrow AI to Singularity, provides insight into how AI works, and the limitations of contemporary algorithms.

3.1 What is an AI model?

The notion of enhancing physical and mental capabilities has been a part of myths and stories across civilizations. As humans, we are in search of an advantage, seeking ways to simplify daily life and comprehend what is around us. Ancient sagas depict intelligent beings bearing extraordinary gifts. Prometheus is said to have brought fire to humanity, thereby establishing knowledge, technology, and civilization. The myth of Hephaestus, the Greek god of blacksmiths and craftsmen, says he created automatons to do his labor. However, tales of artificial intelligence were mere fantasies until the 18th century, when innovations in mathematics, the scientific method, and logical reasoning charted a course towards computing (Coward 2017).

Computing is fundamentally rooted in logical reasoning, representing a systematic and efficient model of human cognition. At its essence, it involves a progression of arguments that are methodically designed to culminate in a conclusion via rigorous reasoning. These arguments are founded on a binary true/false framework, which was first developed by mathematicians such as George Boole (Boole 1854). Charles Babbage and Ada Lovelace utilized binary logic in programming their mechanical calculator, and subsequently, it was leveraged for the development of mainframe computers in the mid-20th century. Currently, binary logic remains an integral aspect of machine language, utilized by the neural networks hosting AI models. Beyond the utilization of binary code, the shared aspect amongst all these technological exemplars lies in their computational approach of modeling the human cognitive process to solve logical problems. As Alan Turing posited, computers are designed to "think" through their capacity to reason systematically and logically (A. M. Turing, Computing Machinery and Intelligence 1950, 433).

With over 1,000 AI-focused scholarly articles published daily (Zhang, et al. 2022), we can find a plethora of research presenting refinements to Artificial Intelligence models (*figure 16*). The phrase "Artificial Intelligence" can conjure images of a computer imbued with the ability to perform human tasks at a superior level. A sort of AI super assistant, an extraordinary accountant, or a car that effortlessly navigates crowded streets. This assumption is not an accurate description of most contemporary AI models. While our computers are demonstrating remarkable human-like abilities, they continue to fall short of functions that can be classified as autonomous, self-conscious, or fully competent (Eisikovits and Feldman 2021). They are adept at mimicking specific human behaviors but lack the independence possessed by humans.

AI is divided into several types of ability, depending on complexity. The category of AI that can perform human-like tasks, such as autonomous driving, is classified as Artificial General Intelligence (AGI). While appealing in theory, and promising in research, researchers have yet to create a reliable AGI model. Most of the AI in our hardware and homes, celebrated in the media, and discussed in this thesis, is

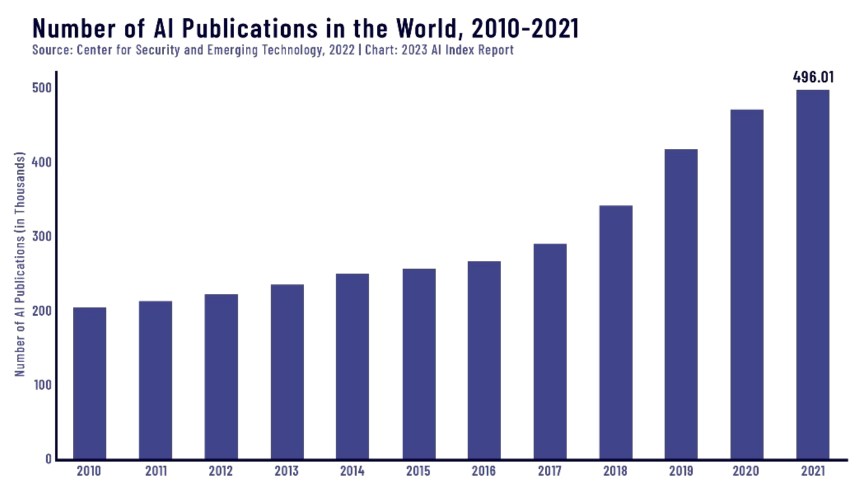


Figure 16 Data source: Center for Security and Emerging Technology, 2022
Chart adapted from: 2023 AI Index Report

classified as Artificial Narrow Intelligence (ANI). ANI applications operate within a limited scope of capability, executing specific functions. While Alexa can turn off your porch light, it is unable to drive a car. Copilot may create a chart for you in Excel, but it will not use the ingredients in your refrigerator to make a recipe. While a diffusion model like DALL·E 2 can create fanciful images, it would be at a loss if prompted to use the Pythagorean theorem to solve the hypotenuse of a triangle. In comparison to previous AI models, the differentiating strength with Artificial Narrow Intelligence is that it first learns how to solve the problem, even if that solution is inelegant. And while it is an inelegant solution, it will repeatedly perform the task, often millions of times, to discover the best result. ANI's capability is derived from their capacity to work at scale within a narrowly defined scope of operation.

Diffusion models like DALL·E -2 and Midjourney appear to output images that display human-like abilities, but upon closer inspection the results often fail to represent the human experience. The shallow representation and incoherent point of view often project the illusion of creativity without an artist's emotional depth. In his essay "Artificial consciousness - Artificial Art", Mike King writes that, "Art and creativity are fundamental features inherent in human intelligence" (King 2002, 152). While AI is built on models that attempt to replicate human intelligence and mirror the human experience, it lacks the human aspect of consciousness. Diffusion models present a thin veneer of intelligence, elements that King identified as awareness, will, perception, thought, memory, creativity, identity, and autonomy. Yet their lack of autonomy and narrow scope poses a risk of relegating the current wave of generative AI to little more than last year's parlor trick, a demonstration of the potential to reproduce our capacity to think but limited by its narrow capability. Predictions of AI's potential, projections presented by both researchers and fantasists, often fall short of initial expectations. Once the novelty has worn off with this iteration of AI, will we discover that the development of AGI is beyond our reach? Evan Puschak said that "Human beings love to write the words, *This will change everything!* Only to shrug a year later when *This changed very little*" (Puschak 2022, 0:24). The goal of creating a model of Artificial General Intelligence is still being pursued in thousands of research labs around the world, but how will they get there? And how will it shape the way we tell stories?

For researchers, the inner workings of these algorithms are largely theoretical. They are unable to prove how they work. As models become increasingly more complex, they are finding it difficult to explain an algorithm's behavior, unable to determine why it generates specific outputs. In describing the inner workings of the chess algorithm AlphaZero's transformative AI, Holden Karnofsky pointed out that the model is poorly understood. "We know that it *works*. But we don't really know *what it's thinking*. As with a human brain, we can mostly only guess at what the different parts of the *digital brain* are doing"

(Karnofsky 2021). These deep neural networks are made up of layers of processing systems trained on human-created data. They are designed to mimic the neural networks of our brains, but also seem to mirror the human behaviors of opaque thinking and inexplicability (Xiang 2022). In his paper, "Unexplainability and Incomprehensibility of Artificial Intelligence", Roman Yampolskiy argues that explainability and comprehensibility are essential criteria when deploying intelligent systems in real-world contexts (Yampolskiy 2019). Users want to comprehend how the decisions are made that directly affect them. If our only reference is an inscrutable algorithm, then Yampolskiy argues that it is impossible to understand the rationale behind failure. Furthermore, if we grow accustomed to accepting AI's answers without an explanation, essentially treating it as an Oracle system, it will become difficult to detect potential wrong or manipulative answers from the system.

For decades, Artificial Intelligence projects have relied on the knowledge and expertise of human engineers to develop systems which are both explicitly designed and easily understandable. Most expert systems were based on decision trees, a logical model of human decision making. The effect of this practice was that both developers and users could naturally assess the accuracy of the output and the process for reaching the answer. However, over the last decade AI projects have shifted from linear decision trees to machine learning systems based on Deep Neural Networks (DNN), at a cost of sacrificing our understanding of exactly how they work. While their inner processes are opaque the appeal is they appear to work. Yampolskiy writes that the continued development of these models indicates developers are satisfied working with this conundrum for as long as they receive the expected outputs. "As long as Big Data and Huge Compute are available, zero human knowledge is required to achieve superhuman performance" (Yampolskiy 2019, 2).

Addressing problems like bias, where the training data's racial and gender bias results in those biases being woven into the fabric of the AI model, becomes difficult to fix when developers do not understand how a system works. Another risk is that the system may be making an unbelievably bad decision, but researchers are unable to intervene because they do not understand its reasoning. This lack of understanding only exacerbates accountability for the responsible development and deployment of AI. As these systems become more complex, with humans less capable of understanding them, there is a call from researchers to begin to focus more on the process of AI and less on the results (Xiang 2022). Yampolskiy notes that hundreds of papers have been published on what has been called eXplainable Artificial Intelligence (Yampolskiy 2019). Unfortunately, this research has led to limited success as researchers point to problems like slow operation, inefficient training or a lack of computing power compared to the opaque AI models. Computer Scientist Jeff Clune notes that, "There are many tasks right now where black box approaches are far and away better than interpretable models" (Xiang 2022). The difference in capability between an eXplainable Artificial Intelligence system and an opaque AI model is often so significant that the opaque version appears to be the only viable solution. Often, researchers are unwilling to work with a less effective tool.

The call for researchers to focus more on the process of AI rather than the output is not going unheeded. A group of researchers believe they may have unraveled one of the mysteries around the inner working of AI language models. In their paper "What Learning Algorithm is In-context Learning? Investigations with Linear Models", the research team, led by Ekin Akyürek, write about a specific behavior with neural sequence models, like Large Language Models (LLM), where they exhibit a remarkable capacity for in-context learning (Akyürek, et al. 2022b). These AI models can, in essence, teach themselves how to understand and process tasks they have not been specifically trained to perform.

LLMs, like GPT-3 and Google's LaMDA, can construct new predictors from sequences of examples in the input without further parameter updates. How the models accomplish this has perplexed researchers and is of particular interest because it ties directly to how an AI model arrives at its intricately detailed output. Usually, to achieve a novel task, machine learning models must be re-trained on fresh data. To re-train the system researchers would be faced with the laborious and time-consuming task of inputting thousands, if not millions, of new data points. Akyürek's team recognized that with in-context learning, the system can learn to reliably perform new tasks from only a few examples, essentially acquiring new skills with no human guidance or intervention. Researchers presented a prompt to the language model, a list of inputs and outputs, and asked it to make predictions about an alternative task it was not explicitly trained for. Understanding how and why these models are capable of retraining presents insights into how language models learn and store information. Akyürek, the lead author of the study, writes that transformer-based in-context models implicitly implement standard learning algorithms by encoding smaller models in their memory, updating the new smaller models as new examples appeared (ibid. 2022b). In an interview with journalist Tatyana Woodall, Akyürek noted that the model is teaching itself something new, "learning from examples on the fly without any parameter update we apply to the model" (Woodall 2023). The implication of the research is the model is not simply copying training data, it is likely building on previous knowledge. He believes this is similar to the cognitive behavior of humans and animals. By tracing the transformer model's ability to write its own machine learning model in its hidden states, the virtual space between the input and output layer of the system, the researchers believe it is empirically possible to see how it mirrors

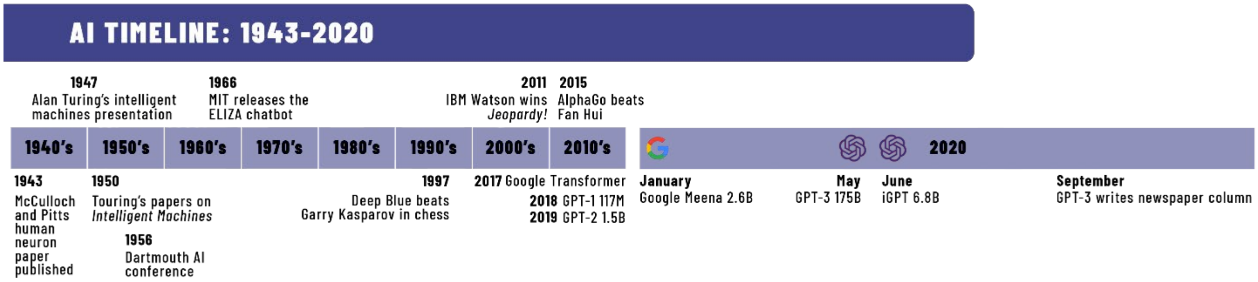


Figure 17 Timeline of AI and language models 1947-2020

human learning. They believe it is creating small hidden state algorithms that are similar to well-known and extensively studied human-created learning algorithms.

3.2 The rise of intelligent machines

Homer's depiction of Hephaestus's automatons as 'thinking machines' is just one illustration of the longstanding literary practice of ascribing human features to machines (McCorduck 2004). While certainly less anthropomorphic than Homer, mathematicians, philosophers, and scientists have also questioned if the human mind is computable. Their search for an understanding of human cognitive mechanics is driven by their desire to algorithmically replicate the process. This pursuit of emulating human thinking has been a foundational concept that frames the development of Artificial Intelligence (Kurt 2018, 2). Like the other inventions, initial efforts to acquire this knowledge were an iterative process, encompassing multiple fields of research. Efforts that did not begin in earnest until the close of the Second World War (*see figure 17*) (Eisikovits and Feldman 2021).

The model for the 'thinking machine' did not come from mathematics, nor was it rooted in philosophy. Instead, it came from the study of physiology. In 1943, neuroscientist Warren McCulloch and logician Walter Pitts modeled the human neuron as a switch that receives input from other neurons and, depending on the total weighted input, is either activated or remains inactive (McCulloch and Pitts 1943). McCulloch and Pitts described a process where the neuron receives an input signal along the outer branches of a dendrite, performs computational switching inside the soma, passes the output through a cable-like structure to the synaptic ends, where the information is passed along to the next neuron (*see figure 18*). The signals in the neuron, called synaptic weights, are binary: either positive (excitatory) or negative (inhibitory). This model was significant because computer scientists realized the depiction of the brain's neural biological switching could be applied as a theoretical approach to the electronic switching in a computer. Both the neuron and computer receive an input, process the signal, and produce an output (Krogh 2008).

Improvements in computing capacity during the late 1940's, led by the post-war governments of Britain and the United States, led both researchers and the public to imagine a world filled with intelligent machines. The term "intelligent machine" was coined by Alan Turing during a lecture to the London Mathematical Society on February 20, 1947, where he discussed development of the British Automatic Computing Engine, or ACE (Copeland 2004). Turing presented his view that this machine held the potential to exhibit intelligent behavior, believing it could display intelligence by learning chess and outplaying human opponents. He suggested that by allowing machines to modify their own commands they could then acquire knowledge from experience, noting that the mechanism of machine intelligence could stem from "[t]he possibility of letting the machine alter its own instructions" (ibid. 2004, 375). As B. Jack Copeland notes in the book, *Essential Turing*, "Artificial Intelligence was not far from Turing's thoughts — he described himself as building *a brain"*. In projecting the computer's potential playing ability, Turing noted it would be rather easy to create a program where it could play a poor game of chess, but again noted "[t]here are indications however that it is possible to make the machine display intelligence" (ibid. 2004, 374). Turing believed that chess, a game favored by mathematicians, where success is often determined by a player's capacity for pattern recognition, was the ideal platform to display the potential of computational intelligence.

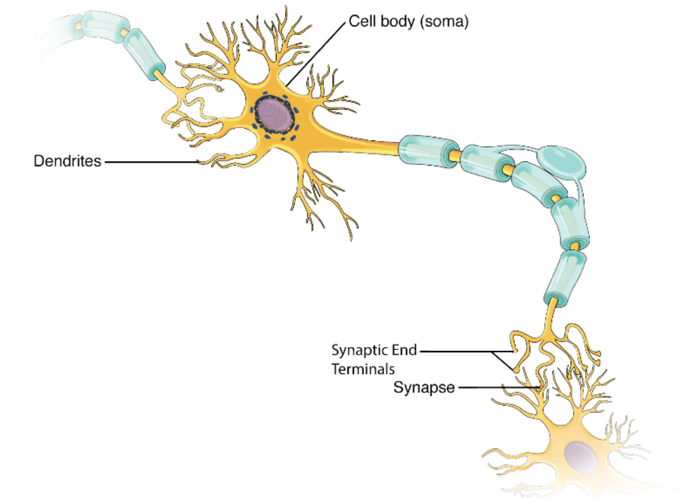


Figure 18 Parts of a Multipolar Neuron,
Adapted from: Oregon State University (Biga, et al. 2022)

The 1940's reports of using a computer to model the human cognitive process spawned fearful stories in the press about these new "electronic brains". Professor Diane Proudfoot, of the University of Canterbury, in New Zealand, writes that newspapers were filled with colorful descriptions of flashing room-sized computers, referring to them as "the controlled monster". The news reports focused on a common fear that developers would be unable to control their inventions, with the implication that it would become "the monster in control". Stories described a popular fear that computers would reduce people to "degenerate serfs", breathlessly predicting that humanity would "perish, victims of their own brain products." Proudfoot noted that, "Turing's response to AI panic was gentle mockery." (Proudfoot, Mocking AI Panic 2015, 2)

This post-war period also saw research in the field of cybernetics, where researchers strove to create practical systems that integrate humans and machines working together to accomplish a functional task (Mindell 1995). For example, an anti-aircraft gun control cybernetic system would integrate human targeting and firing with the machine executing the target tracking and aiming. Researchers were considering how these hybrid human-machine systems might be imagined and designed. In a letter to cyberneticist W. Ross Ashby, Turing expressed a desire to assist Ashby with his research, partially out of frustration with the limitations placed on his research. He wrote that, "[i]n working on the ACE I am more interested in the possibility of producing models of the action of the brain than in the practical applications to computing" (Copeland 2004, 374). He bemoaned that the initial use of the computer was producing output that he described as being devoid of original thinking. But he added that he believed there was nothing in the construction of the system that required it to be used in this sort of predictive manner. "It would be quite possible for the machine to try out variations of behavior and accept or reject them…, and I have been hoping to make the machine do this" (ibid. 2004, 374). Turing understood the importance of the neuron findings of McCulloch and Pitts, observing how his computational model emulated the human brain. He felt that creating a software model within ACE, which afforded the computer to emulate the brain's process of changing neuron circuits with the growth of axons and dendrites within its electronic memory, would facilitate a computational learning process.

By the 1950s, computers were becoming commercially available, and the formal analysis of computational techniques blossomed into the academic field of computer science. During the first eight years of development, from 1947 to 1955, the study of intelligent machines was a small subset within that larger field of computer science, lumped in with mathematicians, psychologists, and electrical engineers. Dartmouth College Assistant Mathematics Professor John McCarthy was editing a book of collected papers, *Automata Studies* (Shannon and McCarthy 1956), and was struck by the emphasis of the findings submitted by researchers. He noted that contributors failed to focus on the computer's potential to possess intelligence beyond simple behaviors (McCorduck 2004). To clarify the prospective intelligence of these machines, in 1955 he organized a small group of researchers to focus on the study of intelligent machines. John McCarthy, Claude E. Shannon, Nathaniel Rochester, and Marvin L. Minsky drafted a 17-page conference proposal outlining a research project where they would spend the following summer hosting a Dartmouth College conference. They planned to gather computer researchers and scientists to study "every aspect of learning or any other feature of intelligence, can in principle, be so precisely described, that a machine can be made to simulate it," naming this new field of study “artificial intelligence” (McCarthy, et al. 2006, 1). They proposed that the group would study how computers use language, form abstractions, unravel how computers learn, and solve problems that were previously reserved for humans. They postulated that a carefully selected group of scientists, working together, would make significant advances in AI. The 1956 Dartmouth conference would be the largest AI-focused gathering to date.

The next summer, participants endeavored to chart a path for future research, with the hope that the path would ultimately shape all phases of AI development. Their efforts included the initiation of symbolic methods and a presentation about the first working AI program, the Logic Theorist (McCorduck 2004). While this new model would ultimately dominate AI research for the next ten years, it appears its significance was unclear to everyone but its developers (ibid. 2004). Conference participants also noted the growth of the mainframe computer’s electronic capacity, observing that computing functionality was doubling approximately every eighteen months (Iku 2021). This observation was made nearly a decade before Gordon Moore, the director of research at Fairchild Semiconductor, posited a similar theory that the number of integrated circuits would double every 18 months (Moore 1965). A theory now referred to as Moore’s law.

When the Dartmouth conference concluded, concrete results were scant. McCarthy recalled that the scientists were rather stubborn about pursuing only the ideas they personally brought to New Hampshire. He was disappointed there was little exchange of ideas between the participants, with some attendees staying for only a few days. Many researchers were reluctant to embrace the term Artificial Intelligence, preferring the term Automata because it appeared more scholarly. Years later Marvin Minsky commented that, “It took another ten years before people could tolerate the idea of AI without thinking that it was funny and impossible” (McCorduck 2004, 117). Still, the conversations at the Dartmouth conference ultimately resulted in the sort of cross-pollination of ideas that McCarthy had envisioned. Ray Solomonoff remembers being particularly taken with an idea of McCarthy’s about the possibilities of expressing intellectual problems in terms of a Turing machine. McCarthy believed that Turing’s hypothetical machine, where a string of symbols goes in to be processed and later a string of other symbols comes out, could be inverted (ibid. 2004, 116). He argued that examining the symbolic output string could present an avenue where you could determine the original input. McCarthy’s 1956 description of inductive inference was an important theoretical steppingstone in the development of Artificial Intelligence (Angluin and Smith 1983).

The researchers who participated in the Dartmouth conference were striving to discover the farthest reaches of computing. They believed computers were more than an adding machine and that their work would be transformational, transmuting the computers into intelligent machines. Still, McCarthy's disappointment over the near-term outcome reflects a persistent problem with artificial intelligence—making machines think, designing computer programs to behave intelligently, would be significantly more difficult than anyone anticipated.

In the 1950s and 1960s, technologists and futurists envisioned a world in which autonomous robots and self-aware artificial intelligence would be ubiquitous. This vision of a machine-driven world has manifested in some contemporary sectors. Robots and intelligent machines are now commonplace in domestic, industrial, and professional settings, performing tasks such as assembling vehicles, engaging in warfare, executing stock market trades, predicting personal preferences, and even vacuuming your house (Sandler 2014). Advancements in AI are occurring at a nearly unfathomable rate, driven by innovations in hardware, training methods, and neural network architecture. We are witnessing progress that appears to be accelerating beyond Moore's Law (Summerlin 2022), reaching a rate of technological innovation that could radically transform all aspects of our life through the merging of human and machine intelligence (Sandler 2014). In 2005 Ray Kurzweil argued that that the rate of technological innovation is increasing exponentially and will result in a period of dramatic transformation (R. Kurzweil, The Singularity is Near 2005). His prediction that we are only decades away from a time where the distinction between human and artificial intelligence will be negligible, with the majority of our intelligence being non-biological, seems a long way from Alan Turing's question, "Can machines think?" (A. M. Turing, Computing Machinery and Intelligence 1950, 433).

3.3 Turbochamp and Sonnets

In June 1949, physicist Douglas Hartree published *Calculating Instruments and Machines* (1950), arguing that while these new electronic computing machines were significant, they should be seen as nothing more than computational engines. His thinking echoed the belief held by many researchers that computers were essentially calculators, lacking the capabilities of an 'electronic brain'. At the same time, Professor Geoffrey Jefferson, a neurosurgeon and Professor at the University of Manchester, was awarded the Lister Medal in recognition of his contributions to surgical science. During his acceptance speech, Jefferson delivered a landmark discussion on computing. Now referred to as the Lister Oration, he outlined his criteria for accepting the premise that 'machine equals brain' (Jefferson 1949). Jefferson argued that to attribute the capacity of thinking to a computer it must be capable of both writing a sonnet and feeling its significance.

At the time of Jefferson's speech in the summer of 1949, Touring had already invested 13 years exploring the theoretical potential of intelligent machines. In 1936, at the age of 23, he published "On Computable Numbers, with an Application to the Entscheidungsproblem [Decision Problem]" (A. Turing 1936), a paper that academic biographer Jack Copeland described as "his most important theoretical work" (Copeland 2004, 1). In the article Turing proposed what he referred to as a digital computing machine, a logical algorithm running on an intelligent machine. This machine was not limited to solving complex mathematical formulas. Instead, it was designed with the capacity to store what it had learned and apply that knowledge when presented with subsequent problems. Now referred to simply as the universal Turing machine, his prescient paper defined the architecture leveraged in modern computers. In 1938, he earned his Ph.D. at Princeton University, discussing the implications of Gödel's incompleteness result by presenting a new analysis of mathematical reasoning (ibid. 2004, 126). Turing returned to England just before the outbreak of the Second World War, working in Bletchley Park, the British government's Code and Cypher School. There he broke the Nazi's Enigma message encryption system and was the principal designer of the 'bombe', a high-speed codebreaking machine (ibid. 2004, 2). When the war concluded, he was recruited to work at London's National Physical Laboratory, designing and developing the same digital computer that he had proposed 10 years earlier. His idea of a universal stored-program computing machine, and his proposal for situated AI, were foundational building blocks in the race to create Artificial Intelligence in both the United States and the UK (ibid. 2204, 439).

In a pioneering but unpublished 1945 paper, Turing first proposed what he called, "Intelligent Machinery" (McCorduck 2004, xxvi). This work would eventually resurface three years later in his 1948 National Physical Laboratory report also titled "Intelligent Machinery". Turing's manifesto about what he called 'machine intelligence', formalized the methods of this new field and included proposals for connectionist-style neural simulation (Copeland 2004). Only a few months earlier Turing had delivered a glimpse of this new field of study in his 1947 London Mathematical Society lecture, uttering what is believed to be the earliest public mention of the phrase "computer intelligence" (ibid. 2004, 354) The challenge he faced was how to demonstrate his machine's problem-solving ability. He landed on the mathematician's favorite game, chess.

At Bletchley Park, Turing had enthusiastically discussed the mechanization of chess with Donald Michie and other colleagues. Michie, a noted codebreaker, reflected that Turing spoke often about the potential of computing machines to solve problems through the application of heuristics to explore the space of potential solutions. Between 1948 and 1950 Turing applied this logic to write his code, iterating and refining the algorithm, eventually naming it "Turbochamp" (Peiravi 2015). He based his program on a method called "generate-and-test". To determine the next move in the game potential solutions would be generated by means of a guided search. These potential solutions would be tested to determine if any of the options are an actual solution. Turing followed this logic because he hypothesized that human decision making follows the same pattern, adding "intellectual activity consists mainly of various kinds of search" (Copeland 2004, 354).

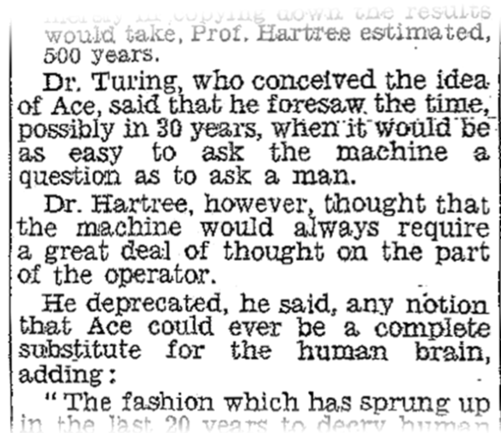
With the National Physical Laboratory's Automatic Computing Engine unavailable, Turing attempted to implement his "Turbochamp" program on Manchester University's Ferranti Mark I computer, without success. He chose instead to demonstrate the algorithm's capabilities without using a computer, with Turing acting as a human Central Processing Unit, running the program with pencil and paper. He challenged his friend and colleague Alick Glennie to a match. When it was Turing's turn to make a move he would consult the algorithm, using its "logic" to decide which pieces to move, and where. It was a slow and laborious process, with each move taking between 15-to-30 minutes. Each time it was his turn he would need to work through the algorithm, analyzing each move, following the steps outlined in his program. While the outcome of the game demonstrated that the "Turbochamp" algorithm was fully capable of playing against a human, it did not demonstrate that it could win. Glennie defeated Turing in just 29 moves (Howard 1997).

Turing's work on the chess algorithm advanced his effort in developing an intellectual scaffolding that could determine if a computer could be programmed to display abilities rivaling human intelligence. He theorized that if a human was unaware whether their opponent was human or computer, and if through play they were unable to determine if it was an imitation or human, then one could argue the computer was displaying intelligence. Many were skeptical of his effort. Since the development of the Babbage and Lovelace Difference Engine, mathematicians and scholars believed the potential of these intelligent machines was limited to mathematical calculations. It was Turing's goal to establish a framework of machine intelligence that would stand up to their scrutiny, disproving what had become known as the Lovelace–Hartree thesis (Gonçalves 2022).

English mathematician and physicist Douglas Rayner Hartree doubted that machines could possess free will (Proudfoot 2020). In a Times of London article, he stated that "These machines can only do precisely what they are instructed to do by the operators who set them up" (D. R. Hartree 1946). Diane Proudfoot writes that this claim is a direct reference to Ada Lovelace in her 1843 response to Luigi Menabrea's Analytical Engine proposal. Lovelace had written that one needed to guard against exaggerating the power of the Analytical Engine. Proudfoot quotes Lovelace's assertion that, "The

Analytical Engine has no pretensions whatever to originate anything. It can do whatever we know how to order it to perform" (Proudfoot 2020, 508). For the next 100 years many mathematicians, physicists and engineers, echoed Lovelace's viewpoint.

Hartree was a leader of a core group of influential scientists who believed Lovelace's assertion was correct. A central figure in the debate over the capabilities of computers, he often voiced his skepticism about a computer's ability to display intelligence in the press. In a 1946 letter to the Times of London he wrote that using the term "electronic brain" was misleading because, "this is just what it cannot do," adding that the term "would suggest to the layman that equipment of this kind could "think for itself" (D. R. Hartree 1946). In a 1946 London Daily Telegraph article announcing that the ACE computer would be used in the study of aerodynamics *(see fig. 19),* Turing mentioned that he saw a time when one could ask the machine a question in the same manner one would address a human. Hartree countered Turing's statement, depreciating any notion that ACE could be a substitute for the human brain (The Daily Telegraph 1946). Still, Touring stood by his belief that the game of chess could demonstrate a computer's ability to learn from experience, and that a successful demonstration would be a direct response to the Lovelace–Hartree thesis.


would take, Prof. Hartree estimated, 500 years.

Dr. Turing, who conceived the idea of Ace, said that he foresaw the time, possibly in 30 years, when it would be as easy to ask the machine a question as to ask a man.

Dr. Hartree, however, thought that the machine would always require a great deal of thought on the part of the operator.

He deprecated, he said, any notion that Ace could ever be a complete substitute for the human brain, adding:

"The fashion which has sprung up

Figure 19 Daily Telegraph, November 8, 1946

When Geoffrey Jefferson delivered his Lister Oration in June of 1949, it was set against this backdrop of competing beliefs. Jefferson was one of many who saw the potential of computing as little more than an efficient way to rapidly calculate solutions to complicated math problems. Nearing the end of his oration he focused on human speech as the highest form of intelligence. Framed by French philosopher René Descartes's viva voce examination for distinguishing humans from machines, Jefferson proposed that to demonstrate thinking, a machine should not only be able to write a sonnet but should be able to discuss its meaning and feel its significance (Gonçalves 2022).

The Times of London reported on Jefferson's speech the next day, reaching out to Turing for a response *(see fig. 20).* Turing offered a pointed rebuttal to Jefferson, offering that "I do not think you can even draw the line about sonnets, though the comparison is perhaps a little bit unfair because a sonnet written by a machine will be better appreciated by another machine"" (The London Times 1949).

CALCULUS TO SONNET

Mr. Turing said yesterday: " This is only a foretaste of what is to come, and only the shadow of what is going to be. We have to have some experience with the machine before we really know its capabilities. It may take years before we settle down to the new possibilities, but I do not see why it should not enter any one of the fields normally covered by the human intellect, and eventually compete on equal terms.

" I do not think you can even draw the line about sonnets, though the comparison is perhaps a little bit unfair because a sonnet written by a machine will be better appreciated by another machine."

Mr. Turing added that the university was really interested in the investigation of the possibilities of machines for their own sake. Their research would be directed to finding the degree of intellectual activity of which a machine was capable, and to what extent it could think for itself.

News of the experiments was disclosed by Professor Jefferson in the Lister oration reported in *The Times* yesterday.

Figure 20 The London Times, June 11, 1949 p.2

While Turing had responded to Jefferson's challenge with wit, it appears the speech may have influenced his choice of model for demonstrating machine intelligence. After his 3-year difficulty developing a working software program, Turing abandoned chess in favor of a test of conversation.

In 1950, Turing published perhaps his most widely known paper, "Computing Machinery and Intelligence" (1950). He begins his argument with a provocation, posing a succinct question that encapsulates the entire controversy surrounding machine intelligence, "Can machines think?" He outlines and defends his new test of machine intelligence, a conversation model he calls "The Imitation Game," and then focuses on potential philosophical, theological, and logical opposition to his question. Turing's paper is a direct response to the mind-machine controversy that had spilled from academia into the London papers. Led by Jefferson, Hartree, and mathematician Michael Polanyi, conservatives were attempting to draw a boundary around Turing's view of machine intelligence. Turing was hoping the Imitation Game would move the discussion beyond those imposed boundaries by creating a demonstration that was both accessible by the public and irrefutable by scholarly examination. Mathematician Jonathan Swinton writes in his Turing biography, *Alan Turing's Manchester*, that it was "Jefferson's obtuseness that provoked Turing into developing this vivid image" (Swinton 2022, 93). Turing's paper specifically references Jefferson's Lister Oration, refuting the position that for a machine to be intelligent it must not only perform intellectually like a human but feel like a human. Turing cuttingly responds that no computer can feel pleasure or grief, know the charms of sex, suffer the misery of mistakes, nor show petulance when it is unable to have its own way (A. M. Turing 1950, 446).

The focus of the paper is The Imitation Game, now almost universally referred to as "The Turing Test." A limited form of the test had first appeared under the heading 'Intelligence as an emotional concept' in the last two paragraphs of Turing's 1948 "Intelligent Machinery" paper (A. M. Turing 1996, 23). Two years later he refined the model. While initially described in the 1950 paper as a game involving men and women, with the goal being to identify one of the participant's gender, the game was eventually revised to involve three different participants: a computer, a human ‘foil’, and a human interrogator.

The interrogator, not knowing which participant is human and which is the computer, asks questions of one of the participants. The other participant’s task, the foil, is to help the interrogator make the correct identification. The goal of the game is for the interrogator to determine if the participant they are interviewing is human or computer. There are no restrictions on the questions posed by the interrogator. They can choose to be as obtuse, wide-ranging, or specific as needed. The participant being interviewed is permitted to say anything to confuse the interrogator. Communication is via text, words written via a keyboard or screen. Successfully confusing the interrogator to where an incorrect choice is made, human or computer, is Turing’s proposed criterion for ‘‘thinking’’ (A. M. Turing 1950).

The purpose of Turing’s work, and the imagination game itself, are still debated. Since its publication, it has received considerable attention from philosophers, computer scientists, psychologists, and others (Copeland 2004, 436). In his paper, “Can machines think? The controversy that led to the Turing test”, Bernardo Gonçalves writes that the paper is seen as complex and multilayered, with most of the key questions still unresolved (Gonçalves 2022, 1). Yale professor Drew McDermott wrote that, “considering the importance Turing’s Imitation Game has assumed in the philosophy-of-mind literature of the last fifty years, it is a pity he was not clearer about what the game was exactly” (McDermott 2014, 2). The paper’s accessibility for a general reader masks Turing’s complex and multilayered thinking. Gonçalves notes that its ambiguity, and often contradictory nature, makes definitive scientific and philosophical interpretation difficult. Interpreters even disagree on whether Turing proposed a definite experiment for determining machine intelligence. There is a widely held view that the imagination game is a behaviorist test of thinking, a measurement of intelligence judging the ability to produce coherent verbal responses to verbal stimuli. Or, as Proudfoot defines this behaviorist view in her paper, “Rethinking Turing’s Test and the Philosophical Implications”, “the philosophical thesis underlying the test is, roughly speaking: if the computer behaves as if it is intelligent, then it is intelligent. If it walks like a duck and quacks like a duck, it just is a duck” (Proudfoot 2020, 1). Proudfoot then continues that she believes this standard interpretation is mistaken. She argues that Turing did not propose a test for intelligence, instead Turing was making a technological prediction. She then cites Daniel Dennett’s belief that the main function of the paper was to provide a unifying framework for discussing and refuting common arguments against the possibility of intelligent machines (ibid. 2005, 5).

On initial reading, the paper appears to frame machines as a species in direct opposition to the human species. That interpretation would also be mistaken. Turing's test is about language performance, the ability to demonstrate intellectual thinking through words. He writes that the model is not a test of human performance, specifically noting that the test draws a "sharp line between the physical and the intellectual capacities of a man" (A. M. Turing 1950, 434). The Imitation Game does not compare a computer's capabilities with our ability to do daily tasks -- like shopping for groceries, watching the sun set, or knowing and executing the nuanced task of pruning your roses. Turing's test is not focused on testing a computer's sensorimotor capacity.

Turing not only responded directly to Jefferson's position, but also addressed the Lovelace–Hartree thesis. He offers a forthright response to Lovelace's assertion that machines are unable to self-originate, interpreted by many as machines can never "take us by surprise" (ibid. 1950, 450). He observes that, "Machines take me by surprise with great frequency," adding that he does not expect his response will silence his critics, anticipating that they would assume "such surprises are due to some creative mental act on my part, and reflect no credit on the machine" (ibid. 1950, 450). Turing's argument here feels a little cavalier, dismissing the argument that computers are predisposed to only execute the task for which they have been programmed. Lovelace's assertion is a machine possesses only the capacity to execute what it has been tasked to perform. It has neither the agency nor free will to genuinely surprise us. While a computer can generate an unexpected result, our surprise is often rooted in a lack of understanding about the program and how in actuality it will interpret our request. Users often experience surprising results when they have made an error inputting the prompt or data, and there is the risk of an occasional hardware component failure. After all, the term "bug" is believed to have been coined when a moth was discovered in a relay of Harvard's Mark II computer (Cohen 1994). While it is true that the output of generative AI often creates a sense of wonder, a feeling of surprise over a seemingly random result to our prompt, a genuine surprise would be if we asked an application like DALL·E for a drawing of a cat on the moon and instead it delivered a ham sandwich. Not a drawing of a ham sandwich, but an actual ham sandwich. That would be a surprise. When we ask for a drawing of a cat and the generative algorithm creates four variations on the theme, some better than others, with the choices surpassing what we can personally draw, our feeling of surprise is predicated on its ability to execute our request. While we may not have anticipated the algorithm would generate the specific iterations presented on screen, and the process does feel a bit like a game of Chance where the user feels as if they are rolling the dice, one should not be surprised that it is doing what we asked of it.

Despite holding the belief that computers only execute the task for which they have been programmed, advances in neural networking and natural language models are starting to undermine the gravitas of this argument. The Imitation Game is a test of language, and as humans we originate something each time we engage in conversation. If we use the framework of origination for defining intelligence (Fillmore 1976), then each time we express our thoughts and ideas through words, producing sentences in our natural language that are appropriate within the conversation's context, then the act can be considered intelligent. Artificial Intelligence is beginning to demonstrate a similar ability. All within the context of an appropriate word choice that emulates contextual meaning. Both the human and AI conversations are examples of origination, of creating a string of words that express an idea, point of view, or emotion. Because of this shared surface level demonstration of origination, the continued application of Turing's test as a referential benchmark for evaluating AI creativity and intelligence is likely to continue.

For Turing, the opening question whether machines can think autonomously was eventually qualified as, "too meaningless to deserve discussion" (A. M. Turing 1950, 442). However, if one narrows the question to ask if a computer can effectively play, and win, at the Imitation Game we may be nearing the point in history where we see demonstrations via GPT models that meet his initial definition of machine intelligence. Clearly computers are performing many tasks that we previously believed required some form of human sagacity. Yet the relevance of the actual test is likely viewed by researchers and developers as nothing more than nostalgia. It does not appear that any research is focused on building an AI model that could pass the test. The study of machine learning and cognitive science is like other scientific research fields. The goal is to prove empirically that the results of a theory will stand up to the scrutiny of peer review. If a model were to be considered as passing the Turing Test with certainty it would need to perform reliably at demonstrating the performance of a real human being. Its capacity to interact and engage will need to be indistinguishable. In reviewing recent AI papers, ongoing research work in Artificial Intelligence appears to have no direct relationship to the Imitation Game. The attention is on building a robust understanding of computational systems of intelligence, to tease out a predictable synthesis of intellect as an algorithmic formula, to create machines that usefully surpass or extend human mental abilities. For developers, imitating human conversation in a parlor game is likely to be seen as contributing little to the ambition of creating genuine machine intelligence.

Nearly 75 years have passed since the publishing of "Computing Machinery and Intelligence" (1950). It reflects one aspect of Turing's research and thinking across a remarkable breadth of topics—

mathematical logic and the foundations of mathematics, computer design, mechanical methods in mathematics, cryptanalysis, the nature of intelligence and mind, and the mechanisms of biological growth. British mathematician Max Newman described Turing's overarching enquiry as "the extent and the limitations of mechanistic explanations" (Newman 1955). In the 1950 paper Turing contends that computers would eventually be created that would simulate human behavior and cognitive thinking. He acknowledged that these machines would make occasional mistakes, but that they would also have the potential to create new "statements", likening the output to the human mind (A. M. Turing 1996, 256). He had specifically chosen experiential learning as the best approach to achieve machine intelligence, writing,

> If the machine were able in some way to 'learn by experience' it would be much more impressive. If this were the case there seems to be no real reason why one should not start from a comparatively simple machine, and, by subjecting it to a suitable range of 'experience' transform it into one which was much more elaborate, and was able to deal with a far greater range of contingencies (A. M. Turing 1996, 257).

In "Human versus mechanical intelligence" (2002), Robin Gandy argues that Turing's paper was not targeted as a philosophical treatise, and instead it was produced as propaganda. While Gandy's word choice is somewhat extreme, he believed that Turing was prodding philosophers, mathematicians, and scientists to recognize computers were more than calculators, that they were capable of intelligent behavior. As a friend of Turing, Gandy recounted how Turing wrote the paper quickly and with joy, reading passages aloud with a smile and "sometimes with a giggle" (Gandy 2002, 125). Gandy writes that the continued scrutiny, application and much of the analysis of "Computing Machinery and Intelligence" loads it "with more significance than it was intended to bear" (ibid. 2002, 125).

3.4 Narrow AI

One of the topics discussed by attendees of the 1956 Dartmouth AI conference was a proposed algorithmic model that would leverage an architecture of symbolic methods (McCarthy, et al. 2006). This model was an alternative to one that more closely modeled a McCulloch-Pitts neuron-like system of computation (Copeland 2004). The neuron model was described as 'connectionist', where learning was facilitated through the creation and modification of connections between real or simulated neurons (ibid. 2004), where the symbolic model was an iteration of Turing's intelligent machine, with a string of symbols entered and processed, and the computer outputting a resultant string of symbols (Russell and Norvig 2003). It is uncertain if the conference discussion was the catalyst, but shortly thereafter the

Symbolic AI model gained significantly wider adoption than the neuron model. By the 1960s, research using symbolic models had progressed to a point where developers concluded that this approach was the most promising path for producing a system capable of demonstrating artificial general intelligence (ibid. 2003).

Symbolic Artificial Intelligence refers to a collection of programming approaches that are based on high-level, human-readable symbolic representations of formulas, logic, and search (Garnelo and Shanahan 2019). The next thirty years saw a progression of development with this approach, with researchers emphasizing tools that leveraged logic programming. As is evidenced throughout the advancement of AI, there were debates among researchers about the most promising method for implementing Symbolic AI's logic programming. By the late 1960's, developers considered whether declarative or procedural representations of logic would be the most effective models. 1969 saw the release of Planner, the first application to successfully implement the proceduralist paradigm (Hewitt 1969). Three years later the proceduralist philosophy advanced with the release of Micro-Planner. Likely the most influential implementation of Planner, it employed a natural language model that had been developed by Terry Winograd (Winograd 1972). Winograd believed that all etymological aspects needed to be integrated for a successful model to have a working understanding of a language, including syntax, semantics, and inference (ibid. 1972) In his paper, "Understanding Natural Language" (1972), he describes how Micro-Planner was capable of recognizing grammar, semantics, and could execute general problem solving. What is likely a symptom of the inherent limitations of Symbolic AI, Winograd's approach assumed that a computer must be trained on the specific subject it would be discussing. Therefore, the program provided a detailed model of a particular domain, ultimately limiting its capabilities. The application could enter into a dialog with a user, using a using a model of its own cognitive logic structure to remember and discuss its plans. In his paper he outlined that the knowledge system was built on procedures, rather than rules tables or a list of patterns. He held that developing these special procedural representations for syntax, semantics, and inference, would lead to an improvement in the model's flexibility and power (ibid. 1972).

The race to create intelligent applications like Micro-Planner had begun in earnest in the 1950s, but the development of Artificial Intelligence has not followed a smooth trajectory. It has progressed in fits and starts, with each surge prompted by a technological development, only to see subsequent advancement slow as the capabilities of existing hardware were exhausted. The 1980s ushered in promising applications, such as knowledge-based expert systems. The success of these applications created an expectation among institutions and developers that they could capture both corporate expertise and investment (Russell and Norvig 2003). By the end of the 1980s symbolic AI had become the dominant paradigm of research and implementation, with a 30-year history of success leading to unrealistic expectations. Development was booming, but suddenly innovation was running headlong into the inherent limitations of both hardware and the symbolic model. As the 1980's was coming to a close, advancements in AI were slowing and the economy faltered. The result was a corresponding reduction in the financial support required for research. What unfolded was a chain-reaction of pessimism in the AI community, which was picked up by the press and may have contributed to a severe cutback in funding, ultimately leading to the curtailment of serious research (Crevier 1993).

In 1985 Marvin Minsky warned the research community that enthusiasm for AI had spiraled out of control. He argued that current models of expert systems were rooted in twenty-year-old programming methods and that they had become more powerful because computers had become faster. In his book, *AI: The Tumultuous Search for Artificial Intelligence* (1993), author Daniel Crevier reports Minsky commenting that by the mid-80s, developments in AI were the result of implementing out-of-date methods that were no smarter than their predecessors, "they just made stupid mistakes faster" (ibid. 1993, 203). In 1984 Minsky coined a phrase to describe an approaching decline in Artificial Intelligence research that he was convinced would follow the current period of unbridled exuberance: the AI "Nuclear Winter" (ibid. 1993) *(see fig. 21)*. For developers, the seemingly irreparable technical problems were manifold. They were facing difficulties with system knowledge acquisition, managing large databases, and they were unable to solve problems associated with out-of-domain system

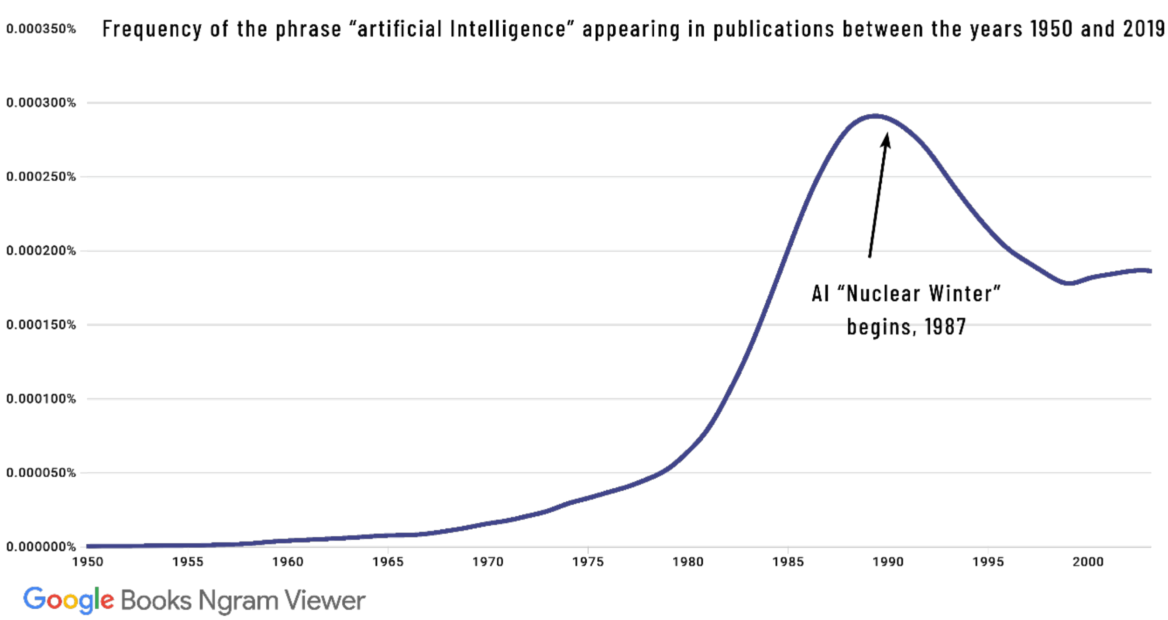


Figure 21 Google Books Ngram Viewer, Sourced February 27, 2023

connectivity (Kautz 2020). The challenges appeared insurmountable with existing technology and the application of the symbolic model. As funding dried up, innovation slowed.

When McCulloch and Pitts first described the neuron as a human switch, they believed that if one could model the human nervous system, then one could potentially replicate the behavior of the brain (Maeda 2022). The theory is the cell acts as a function which receives input via the dendrites, generating an output via the axon terminals (*see fig. 18*). The research into neural networks was focused on finding a way to represent the neuron's function mathematically and then connect a series of these neuron replicators together in a useful way (Jordan 2017). The theory is that the cell acts as a function which receives input via the dendrites, generating an output via the axon terminals. The research into neural networks was focused on finding a way to represent the neuron's function mathematically and then connect a series of these neuron replicators together in a useful way (ibid. 2007). In 1957, American psychologist Frank Rosenblatt advanced the thinking that neural modeling was a path for creating machine intelligence. In his paper, "The perceptron: A probabilistic model for information storage and organization in the brain" (1957), Rosenblatt explored how a biological system, like a human, senses the physical world, how that information is stored, and how memory influences behavior. He proposed a path for creating a neural algorithm model that learns by seeing and remembering, referring to his theoretical nerve net as "perceptron" (Rosenblatt 1957, 387). Like many neuron models, his conduit for emulating intelligence was 'connectionist', where learning is facilitated through the creation and modification of connections between simulated neurons.

In 1957 symbolic logic was the most widely accepted model for developing machine intelligence. Rosenblatt disagreed with that line of thinking, noting that while the development of symbolic logic, digital computers, and switching theory had impressed many theorists with the functional similarity between a neuron and the simple on-off units that make up a computer, he did not believe they were capable of Artificial Intelligence. He bitingly observed, "Unfortunately, the language of symbolic logic and Boolean algebra is less well suited for such investigations" (Rosenblatt 1957, 387).

Four years later Rosenblatt followed up his initial perceptron work with a 626-page report produced by the Cornell Aeronautical Laboratory titled, "Principles of neurodynamics: Perceptrons and the theory of brain mechanisms" (1961). The paper was widely influential, with some US research groups choosing to adopt his approach over symbolic logic (Copeland 2004). Its purpose was to refine his theoretical system of neural modeling. He believed his model explained the psychological functioning of a brain within the known laws of physics and mathematics, and within the known facts of neuroanatomy and physiology (Rosenblatt 1961). Curiously, in his preface he adds a touch of whimsey to frame the public's response to his earlier work. He describes how "perceptron" was intended to be used as a term for classifying theoretical nerve nets, observing that there was an unfortunate tendency for people to assume that a perceptron was a piece of hardware. He laments that people are unable to "suppress their natural urge to capitalize the *P*" (ibid. 1961, vii). Perhaps the most amusing section, from a scholarly report that otherwise conveys a tone that approaches the funerial, is his disappointment over the media's handling of the first public announcement about his work. He describes journalists who exhibited "the exuberance and sense of discretion of a pack of happy bloodhounds" (ibid. 1961, vii). He concludes by expressing his displeasure over headlines that accompanied the news reports, finding one example that appeared in Tulsa's *Oklahoma Times* particularly galling; "Frankenstein Monster Designed by Navy Robot That Thinks". He added that the story was "hardly designed to inspire scientific confidence" (ibid. 1961, vii).

Rosenblatt's work, and the perceptron model, eventually acquired a certain amount of notoriety within the AI community. Many admired the simplicity of the system, but in her book, *Machines Who Think* (2004), Pamela McCorduck describes another reason for his growing fame. It was Frank Rosenblatt himself. She writes that researchers remember Rosenblatt as prone to uttering extravagant statements about the performance of his machine. She quotes one scientist stating that "He was a press agent's dream, a real medicine man. To hear him tell it, the perceptron was capable of fantastic things" (McCorduck 2004, 105). While Rosenblatt may have been prone to hyperbole, his research proved that neuron network models have properties similar to the brain: they can perform sophisticated pattern recognition, and they can function even if some of the neurons are destroyed (Krogh 2008). His demonstration showed that that the perceptron model could solve a limited class of linearly separable problems. Because this line of thinking went against the mainstream of study, a stream populated by the symbolic model, and despite his initial success, by the late 60s research activity in neural processing diminished. Although it did not disappear entirely.

In 2012, neural modeling realized a renewed surge of interest. Even though the capacity for computation was increasing, academics observed the additional power was insufficient when addressing difficult problems using computers developed for symbolic processing. This inherent limitation led researchers to explore alternative paths, eventually concluding that neural modeling had the potential to apply more computational cycles to their algorithms. More cycles meant more math, which meant bigger programs running on outdated networks. The computational requirements of neural models resulted in systems struggling to meet the increased demands.

For any neural network the training phase of deep learning is the most resource-intensive task. This is particularly true when developing an algorithm like natural language processing because the phase demands the most significant number of computational cycles. When developing an algorithm, developers first test with a smaller data set. Usually, these tests can be performed locally on a desktop computer or local network. Then, when training their model with large data sets by operating their algorithm at scale, developers often encounter limitations in the number of computations that can be processed. The consequence in this cap is the frequency of digital computations performed slowed significantly, resulting in processing lag. Tasks that previously took minutes when training with the smaller sets may now take hours, or weeks.

Current AI models are designed to output data that replicates human decision making. To demonstrate this skill, it must first learn how to perform the task. To understand the process required to perform the task, it must learn the necessary steps. It discovers those steps through pattern recognition. A Narrow AI model is trained by setting a goal, defining the parameters, and allowing it to discover the required steps via trial and error. This is why the learning phase is so resource intensive. Before beginning this phase, a teachable neural network model starts with random weight (W) and bias (b) values. While conceptually abstract, when an algorithm is looking for patterns in the data, weights can be thought of as an indicator of the connection's strength. Their purpose is to affect the amount of change in the data's value before it is output. When the weight value is low it does not change the input's original value. Alternatively, a larger weight value will result in a more significant change in the output. As training continues, both parameters are adjusted toward the desired values and the correct output.

During the training phase, a neural network receives inputs which it processes in hidden layers by applying the weights and bias. Researcher Jay Alammar describes the weight's function as the key parameter that transforms the input data (Alammar 2016). In simplified terms, a neural network is a series of nodes that act like our neurons. The perceptron model affords a straightforward way of visualizing a neural unit. Each node is primed to receive data, holding a set of values: input, weight, and bias (*see fig. 22*). The incoming data is multiplied by the value of the weight and then added to the expected bias value. The result of this equation is compared against a threshold value, which sets the status of the node. The resulting status is either "observed" or "passed on" to the next

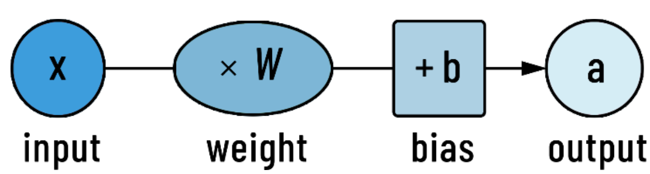
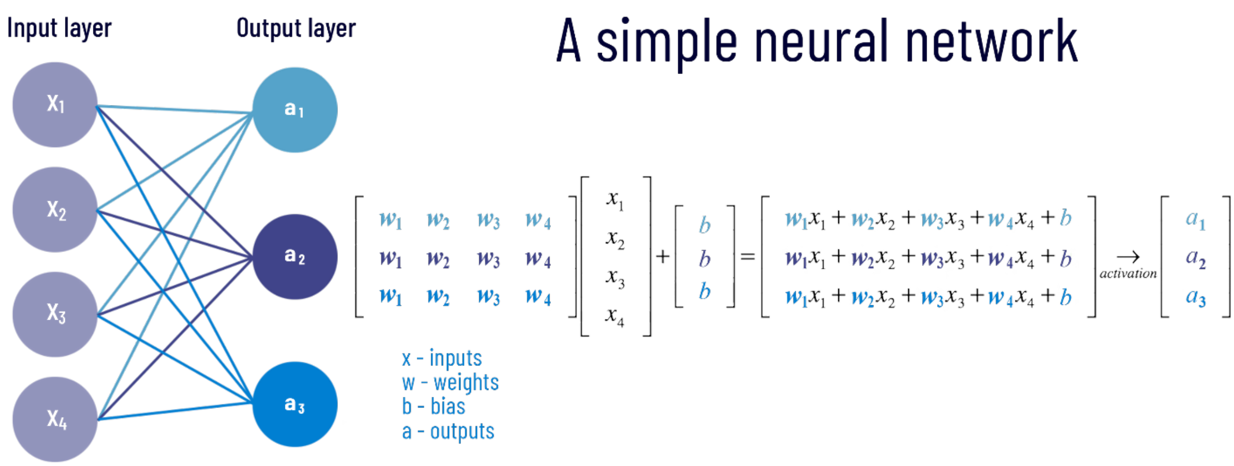


Figure 22 Neural Network node weight operation
Adapted from: A Visual and Interactive Guide to the Basics of Neural Networks (Alammar 2016).

layer of the neural network (ibid. 2016). Because the node is acting as a binary switch one can consider its function to be in a state of on or off, with the computation either flipping the switch's status, or leaving it unchanged. The importance of the weights in deep learning is during training they are adjusted based on what has come before, revealing patterns in the data, and ultimately resulting in better predictions.

The capability of the above model is limited to performing only binary classification. By connecting a collection of models into a matrix an algorithm can then be taught to perform multi-class classification (Jordan 2017). In figure 23 you can begin to visualize how the nodes fit together to form a neural network. This example shows three distinct logistic regression models, each with their own set of parameters. By operating in a matrix, the algorithm can quickly and efficiently perform the required calculations (ibid. 2017).



$$\begin{bmatrix} w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} b \\ b \\ b \end{bmatrix} = \begin{bmatrix} w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \end{bmatrix} \xrightarrow{activation} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

Figure 23 Adapted from: Neural networks: representation (Jordan 2017)

Training an algorithm is often about continuously managing variables, called parameters (Bashiri and Geranmayeh 2011). They can be thought of as a developer's method of tailoring their hypothesis to fit a specific data set. Many early algorithms were leveraging a comparatively small number of parameters, from 100 to 100,000. However, through a decade of technological iteration the number of parameters within these models have grown significantly with a recent publication announcing a Scaling Language model with over 500 billion (Chowdhery, et al. 2022). In clearer terms, the more parameters used in a model, the more computations are required, the more time it will take to run. For training a Deep Learning model the best solution is to execute all the operations in parallel.

While the 2012 researchers were not writing algorithms that leveraged hundreds of millions of parameters, their models were experiencing significant processing lag running on existing computer systems. They were searching for a way to create neural networks that could perform the required calculations quickly and at scale. One solution was incorporating arrays of low-cost Graphics Processing Units (GPU). GPUs proved to be ideally suited for computing the necessary numerical neural calculations through their combination of processing power and relatively low cost (Maeda 2022). If we consider how nodes fit together in a neural network, they are essentially performing matrix multiplications. GPUs are a single-chip processor used for extensive graphical and mathematical computations. They can be devoted to performing calculations via a processor that is separate from the computer's Central Processing Unit (CPU), freeing up the CPU's computing cycles for other tasks. An additional benefit of using a GPU for calculations is they have been designed to perform the accelerated geometric calculations required to render complex graphics. Devoting these proportionately increased number of transistors to arithmetic logic units results in an exponentially amplified application of math. One can consider the processing speed of the CPU as nimble and rapid, like the performance of a Lamborghini. Conversely, the GPU is more analogous to a computational cement mixer. It is not as nimble but can apply significant brute force.

By shifting the required calculations of these ever-growing models from CPU to in-parallel GPU-powered neural networks, developers saw a significant improvement in computational speed. Initially, the GPU solution was an ideal fit for network developers. At the time, the market was flooded with units due to over production. Developers began to create ever more expansive neural networked processing arrays, prompted by not just the new requirements of AI, but new applications like mining for digital currency. One unintended consequence was the number of new networks coming online significantly increased demand, resulting in a spike in the price of available GPUs (Maeda 2022).

As the development of neural networks gathered momentum, other narrowly tasked models were seeing adoption within both the enterprise and consumer space. Called "small models", or "Narrow AI", these models were considered "state of the art" at the beginning of the 2010s (Huang and Grady 2022). AI applications are classified based on their ability to accomplish tasks, with the three classifications progressing according to complexity. The most common classifications are Narrow, General and Super (Coward 2017).

Narrow is the most widely implemented category of AI. Many of these applications are clearly visible in consumer products and services. From Siri on your phone, to the chatbot helping you book a flight, to the recommendations presented on your Netflix screen (Marr 2021). Its widest implementation is found in the products and services that rarely involve direct consumer interaction. This includes smart automation in factories, big data analytics, and credit monitoring. Narrow AI models are constrained to accomplish specific tasks rather than emulating the cognitive abilities of the human brain. They classify data based on their training, with laser-sharp attention to specific goals. Its architecture and operation guarantee the completion of defined tasks, with wide adoption in enterprise due to high performance, speed, and energy efficiency (Coward 2017). Additionally, corporations see benefits because many of these Narrow AI services do not need to run locally on their servers. The barriers to adoption are relatively low because large cloud services like Microsoft Azure, Amazon Web Services, and Google Cloud afford relatively seamless connectivity to data. These "Software as Service" applications are viewed by the multinational providers as an avenue for growing what is an existing position of global market dominance, boosting revenues across their entire portfolio of computing services (Metz and Weise 2023). Cameron Coward notes that implementations of narrow AI monitor your credit card transactions for fraud, reads the numbers you write on your checks, and if "you search for 'sunset' in the pictures on your phone, it is AI vision that finds them" (Coward 2017).

3.5 The race to scale

These global corporations are quietly shepherding the expansion and implementation of Artificial Intelligence through massive capital investments, conducting their development largely outside of public scrutiny. While there is continued research in academic environments, in the last five years the major advancements are being driven by large corporate research institutions. The consequences of this race to scale are likely to be far reaching.

Corporations are in the lead due to the significant cost associated with networked computer processing. Most applications are not running locally on a personal computer, instead they rely on remote servers and the massive neural networks described earlier. What you see on your screen is little more than a visual shell that affords easy-to-use connectivity to central processors. The investments required for high-speed data connections, thousands of computers arrayed on racks, secure warehouse-sized buildings, electric power, climate control systems, specifically designed network architecture for each system, and the personnel required to run the facility, are substantial. Remote server farms are scattered across the globe, positioned largely in remote areas close to an electrical power source, with more coming online to meet an ever-growing demand. Because of these associated costs, the expenses are being borne by corporations who expect a return on their investment.

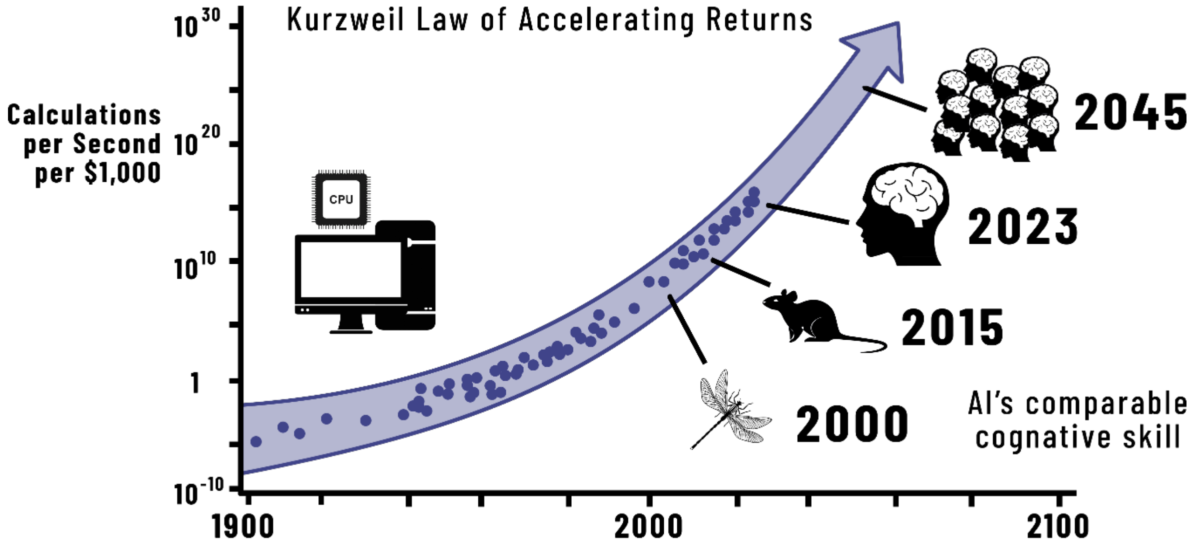


Figure 44 Kurzweil Law of Accelerating Returns, 2014,

The evidence of this race to scale can be seen when you examine how the growth of computational power is driving a corresponding increase in the cognitive capabilities of AI. In his 1999 book, *The Age of Spiritual Machines*, Ray Kurzweil proposed "The Law of Accelerating Returns", where the rate of change in technologies like AI increases exponentially (R. Kurzweil 1999). In 2014 he developed a chart illustrating how as the cost of computing has been reduced, the capabilities of algorithms have increased (*see figure 24*). He plotted the number of calculations per second in relationship to an expenditure of $1,000. For reference, he identified positions on the graph indicating the capability of AI in relationship to the cognitive capacity of biological beings. Kurzweil identifies specific milestones, with the year 2000 representing the capability of a dragonfly, a mouse in 2015, and an estimation that this year these models will have a cognitive capacity comparable to the human brain. He also predicts that with continued exponential growth, by 2045 AI's capacity will equal the sum total of every person on the planet (Maeda 2022). In a research article titled "Spiritual Machine", Kurzweil writes that he believes the world will be transformed to a point where the difference between man and machine blurs, "where the line between humanity and technology fades, and where soul and silicon unite" (R. Kurzweil 1999b). While it can be argued he was overly optimistic when he predicted in 1999 that the next 20 years would bring more change than the previous 100, we have seen exponential growth in both computing power and the capability of algorithms. We are only now beginning to realize the social tension these new capacities are bringing to creativity and the workforce. It is certainly up for debate whether the next 20 years will see the cognitive capacity of Artificial Intelligence morphing into a hypothetical future of singularity, where technology irreversibly grows out of control, "where soul and silicon unite" (ibid. 1999b). We have learned with previous technological innovations that predicting the future should, at best, be considered a crapshoot, with limited perspective on unexpected consequences.

By 2017, researchers were leveraging these new massively parallel networks to develop deeper understandings of natural language models. In a paper titled, "Attention is All You Need", Google Research revealed a new architecture they called transformers (Vaswani, et al. 2017). The significance of this paper was their effort to create a model that ran in parallel and was appreciably easier to train. The process was called "Few-Shot Learning" (FSL), where developers facilitate accurate machine learning using less training data. Previously, creating a machine learning system required gathering massive data sets for use in training the algorithm. Researchers were operating under a belief that increasing a model's requirement to learn by throwing more data at the algorithm would produce a corresponding improvement in the fidelity of the model's output. As previously discussed, the consequence of this approach was an increase in time and cost.

Few-shot learning is significant because the training dataset contains limited information (Dilmegan 2020), thus reducing the time and cost. The model is based on the way humans are capable of discerning differences in objects with only a few examples. Few-shot learning (FSL) can be considered as a meta-learning problem where the model learns how to learn to solve the given problem (Zi, Ghoraie and Prince 2019).

For example, if the algorithm is designed to classify animals from photographs, it could be asked to perform a series of tasks where two pictures of three different animals are presented. The task is for it to classify the images in the query set. This would be considered a 3-way-2-shot classification problem (ibid. 2019). As more images are fed into the model for classification, the performance is measured against an expectation that it will quickly learn how to classify the images. In the example below (*fig. 25*), the algorithm is asked to perform general classification rather than specific, as it has not had the opportunity to learn the difference between a shark and a dolphin, only that they both appear to have similar body shapes. The benefit of a few-shot learning model is they have the flexibility to be customized to specific domains.

Vaswani, et al. were working to build a system that would eliminate the processing caps that were a fundamental constraint of sequential computation. They reported that their Transformer architecture significantly reduced training time, relying entirely on an attention mechanism to develop global dependencies between input and output (Vaswani, et al. 2017). The implications of their work extended far beyond the Google Translate engine they were building, as other researchers saw the

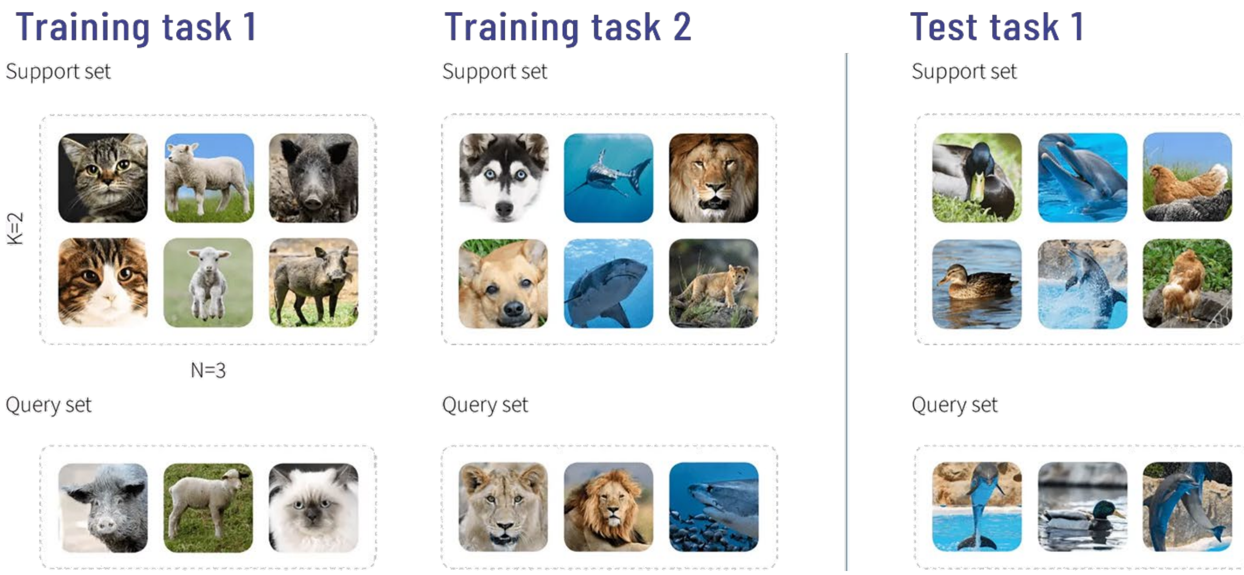


Figure 25 Few-shot learning and meta-learning, 2019 Source: Zi, Ghoraie and Prince

potential to eventually deliver human-level cognition. Matthew Hutson writes that in the five years between 2015 and 2020 the compute used to train the different AI models increased by six orders of magnitude, "and the results surpass human performance benchmarks in handwriting, speech and image recognition, reading comprehension and language understanding" (Hutson 2022).

One of the most significant challenges to developing these AI models, and specifically Natural Language Generation, lies not in modeling the desired output, but the evaluation. Andrew Gibbs-Bravo notes that human evaluation is considered the superior mode of evaluation. The challenge associated with human supervised learning is "it suffers from significant disadvantages, including being expensive, time-consuming, challenging to tune, and lacking reproducibility across experiments and datasets" (Gibbs-Bravo 2019). In his paper, Gibbs-Bravo demonstrates an approach called Reinforcement Learning which in many ways mimics human learning behavior. During the training process the learning agents learn through interacting with their environment, noting the rewards they receive. Gibbs-Bravo believes this approach results in significant improvements in training time and fidelity. Deep Reinforcement Learning was shown to "outperform supervised learning methods in automatically generating paraphrases of input text" (Gibbs-Bravo 2022).

This year a Microsoft research team published a research paper describing their advancement in AI voice cloning they called VALL-E (Wang, et al. 2023). Voice cloning is a process where an AI model listens to samples of a specific individual's voice to learn the characteristics of their speech. Most existing models, like Descript's Overdub service, require approximately 30 minutes of source audio from an individual, usually reading a prepared script that includes required phonemes, intonation, and speech cadence. The significance of the Microsoft research model is VALL-E affords the training to take place with a sample of only 3 seconds (ibid. 2023). Where the Google team was able to significantly improve their language translation model using Few-Shot Learning, and the Gibbs-Bravo research demonstrated the potential of Reinforcement Learning, this application leverages a “Zero-Shot” Learning (ZSL) approach (ibid. 2023). Zero-shot is an approach that affords a model a mechanism for classifying objects from classes that are previously unseen, accomplishing the request without receiving additional training. These types of models are initially pre-trained on a set of classes, in this example a large corpus of various human voice recordings. The model is then asked to generalize to a new data set, the 3-second recording, without additional training. A ZSL transfers knowledge that is gained from the initial training to the task of processing the new request.

The Microsoft team acknowledges the potential risk of misuse, specifically calling out the possibility someone could impersonate an individual without their permission or knowledge (ibid. 2023, 12). The fear of misuse is well founded. Companies, such as AI voice synthesizing ElevenLabs, have released free text-to-speech tools that transform a short vocal sample into a synthetically generated voice. ElevenLabs made news in January of 2023 when their audio synthesis application was used to replicate voices of celebrities saying things they never uttered, such as Emma Watson falsely reciting passages from Adolf Hitler’s “Mein Kampf” (Sellman 2023). The misuse of these AI models is prompting scrutiny in the legal community, with lawyers noting the potential harm to unsuspecting individuals. United States Supreme Court Justice Neil M. Gorsuch indicated in February of 2023 that the legal protections that shield social networks from litigation may not apply to work created by AI (Verma 2023).

The Microsoft team indicates in their paper that the technical capability to build a detection model for determining whether an audio clip was synthesized by VALL-E is feasible. It is unclear from their paper, however, if that process has begun. To their credit, they make it clear they will adhere to Microsoft’s Responsible AI Standard (Microsoft 2023). This is a document that delineates the policies and procedures that are in place for guiding the internal teams that are developing and deploying AI in the company. They refer to it as their “internal playbook”, adding “We are committed to the advancement of AI driven by ethical principles that put people first. Empowering impactful responsible AI practices” (Microsoft 2023).

In 2020, the World Economic Forum (WEF) predicted the next five years would see 85 million workers displaced by Artificial Intelligence. They identified white-collar task-oriented jobs as the likely targets, including data entry clerks, administrative assistants, accounting, and auditing professionals

(World Economic Forum 2020). They anticipated that by 2025, workers who were identified as "redundant" in the workplace would drop from 15.4% of the workforce to 9%, a 6.4% decline. The WEF expressed a belief that while the new 'jobs of tomorrow' would eventually surpass the jobs lost, they observed that job creation was slowing while job destruction accelerated (ibid. 2020). Gartner Insights published a 2020 report estimating that by 2030, 80% of current project management work will be automated, eliminating the discipline by replacing a manager's traditional functions with Artificial Intelligence (Stang 2020). A global survey of corporate executives from the AI platform PEGA also projects AI's probable impact on the number of white-collar jobs. In "The Future of Work*",* they report that 78% of the executives believe their increasing reliance on AI and robots will dramatically reduce the ranks of middle management (PEGA 2022). Responses from a 2020 survey conducted by the Boston Consulting Group reveal that employees holding these mid-level management positions are reaching a breaking point. They do not envision a future where they will have a job (Beauchene and Cunningham 2020). In a 2022 report, "Generative AI: What does it mean in the Enterprise?", the International Data Corporation projects that as AI becomes more mainstream, 97 million new jobs could be created (International Data Corporation 2022). This job creation would potentially offset the WEF's projected loss of 85 million by nearly 15%. It is worth noting, however, the types of jobs IDC identified require significantly different skills than data entry, middle management, and administrative assistance. The 'jobs of tomorrow' they identified will be filled by data scientists, process automation specialists, digital marketing, and strategy experts. The gloomy prognostications and barometers of belief listed above appear to foreshadow a pragmatic understanding, by both workers and executives, that we risk a contemporary "Engels' Pause".

These categories of entirely new job descriptions typify how the disruptions of an emerging technology often eliminate the careers of some, while creating entirely new ones for others. This turmoil is an example of the principle of creative destruction outlined by Austrian political economist Joseph Schumpeter. He offers an important insight into modern economics that an essential element of capitalism is to be found in what he described as the 'perennial gales of creative destruction' (Schumpeter 1942 , 76). Economic change, while promoting innovation and growth, simultaneously tears down existing systems (Schubert 2013). The collateral damage is often the displaced worker. In the 2020 World Economic Forum report described above, they observed that contemporary worker inequality is likely to be exacerbated by a dual impact of technology and an economic recession. They identified lower wage workers, women and younger workers as already feeling the impact, comparing the 2020 economic climate to the Global Financial Crisis of 2008, noting that "the impact today is far more significant and more likely to deepen existing inequalities" (World Economic Forum 2020).

This race to scale Artificial Intelligence into markets, with global organizations pushing the technology ever forward, and workers realizing their livelihood is in peril, illustrates significant growing pains. Applications like Jasper for copywriting, Stability AI for image generation, DoNotPay for legal services, Omnekey for creative marketing, Paige.ai for cancer diagnostics and Mostly.ai for synthetic data sets, point to how quickly industry is deploying a range of game changing innovations (International Data Corporation 2022). Technology reporter Sara Dietschy observes that "all of these AI companies are in a race, a race to the bottom, trying to build better and cheaper options" (Dietschy 2023, 1:04). Writer Maggie Harrison frames the developments with an eye to unanticipated and unseen outcomes, writing "A lot of new tech is getting bottled into consumer-facing products. There is just one problem: neither the products — nor the public — are ready" (Harrison 2022).

3.6 AGI – Artificial General Intelligence

The Artificial Intelligence examples we have considered thus far are typically categorized as Generative Artificial Intelligence. These Narrow AI algorithms, such as ChatGPT and diffusion models, generate novel content using unsupervised and semi-supervised algorithms that are trained on pre-existing data (International Data Corporation 2022). Many of the current applications are focused on creative endeavors, including the generation of audio, computer code, images, text, simulation, and video (McKinsey 2023).

A stated goal within AI research is to step beyond the current neuron modeling of Narrow AI and expand AI's capabilities into a realm where the machine can perform the daily tasks of a human (Kurt 2018). Known as Artificial General Intelligence (AGI), or "Strong AI", this versatile model will have the capacity to think and act like us. The expectation is it will emulate the human mind with the capacity to solve any kind of complex problem (Thompson 2022). To meet this threshold, it will need to demonstrate perceptual abilities, including vision and language processing, as well as cognitive capabilities such as analysis, contextual understanding, and higher order thinking. Strong AI would be capable of actual thought and reasoning and would possess sentience and/or consciousness (Coward 2017). This is the stuff of science fiction and futures studies, with large research organizations like DeepMind, Google and OpenAI actively pursuing development. Author and AI researcher Alan D. Thompson has stated that he believes scientists are "about 30 - 40% of the way to achieving Artificial General Intelligence, where a computer can do anything that a human can do" (Thompson 2022).

Current models of AI sit on a spectrum of what Nir Eisikovits and Dan Feldman have defined as useful machines (Eisikovits and Feldman 2021). Their machines range from simple mechanical tools like the wheel and hammer, to digital tools like Narrow AI that afford a user the ability to set goals with the knowledge that implementation can be left to the machine. What would differentiate today's AI with the development of Artificial General Intelligence would be the latter's ability to function as a human, including the facility to autonomously set its own goals. Eisikovits and Feldman believe that if AGI possesses this ability, it will move AI off the end of their tool spectrum, transforming into something new that will prompt entirely new social anxieties (ibid. 2021). This may be particularly true if the humans in this man/machine equation are unable to identify which of their daily interactions are with an artificial entity. Clifford Nass and Youngme Moon write in their paper, "Machines and mindlessness: Social responses to computers", that when people interact with computers, they perform normative social behaviors and unthinkingly apply social rules (Nass and Moon 2000). A user's expectation is the interaction is occurring with an entity that shares the same values, responses and ethics. Users tend to approach computers as if they are independent from their source of origin. There is research that points to user disassociation between the goals and values of those responsible for creating an application and the functioning behavior of the machine (Sundar and Nass 2000). It brings into focus how computers are being treated as social actors through these interactions (Hertzmann 2018). People unconsciously apply social norms and attitudes to their AI exchanges; consider how they talk to Siri or Alexa. Researchers are applying the theory of Computers Are Social Actors (CASA) to current implementations of AI, with an emphasis on how user behaviors and personality traits change through interaction (Mou and Xu 2017). If the development of AGI is successful, we are likely to observe and experience significant social consequences when users are presented with an artificial general intelligence that matches or exceeds human capabilities.

In his paper, "What is Recursive Self-Improvement?", John Spacey presents a frightening scenario where AGI computers are imbued with the ability to write their own code in repeated cycles of improvement (Spacey 2017). One of the central features of Strong AI is its capacity for unsupervised learning. Theoretically, it would possess the ability to run by itself, set its own goals, upgrade its algorithm, and iterate endless cycles of self-improvement, all without instructions or human agency. As Spacey frames the scenario, "the maker of the AGI will not be in charge of programming all the possibilities or outcomes" (ibid. 2017). This description of a machine's sentient evolution is referred to as 'recursive self-improvement'. Deniz Kurt notes that John McCarthy and Ray Kurzweil have predicted that once developers have achieved a working AGI model, it will quickly surpass human intelligence (Kurt 2018, 18).

Kurzweil has written extensively about his belief computers will soon attain intellectual capacities that surpass those of a human, describing the moment as 'Singularity'. While Kurzweil may be the person currently associated with the concept of Singularity, he is not the person who coined the phrase. In his essay, "The coming technological singularity: How to survive in the post-human era*",* science fiction writer Vernor Vinge described a theoretical threshold where the exponential growth of technology increases until it is uncontrollable, dubbing the moment 'technological singularity' (Vinge 1993). Information scientist Murray Shanahan believes the first person to use the concept of a "singularity" in the technological context was mathematician and computer scientist John von Neumann (Shanahan 2015). Still, it is Kurzweil's extensive writing about the likelihood that AI's exponential growth will lead to superintelligence, with an eventual inflection point of Artificial Superintelligence, that has been defining the discussion.

For Kurzweil, singularity is a point in time where non-biological intelligence matches the range and subtlety of human intelligence (R. Kurzweil 2014). He describes it as the moment where technological change will be so rapid that human life will be irreversibly transformed (ibid. 2014). His vision of Singularity is decidedly agnostic, believing the world will be "neither utopian nor dystopian, this epoch will transform the concepts that we rely on to give meaning to our lives, from our business models to the cycle of human life, including death itself" (ibid. 2014, 207). Kurzweil's description of Singularity can be described as an intellectual treadmill where society is unable to keep up with AI's expanding cognitive capacity. The capability of these machines would match or exceed human movement, thinking and working, operating at a speed where self-improvement and development occurs unseen (Thompson 2022). Kurzweil's conceptualization of singularity also reflects the parallelism between AI research and the studies of neuroscientific and cognitive human intelligence. In his writing he projects that the computational capacity required for AI to emulate the richness, subtlety, and depth of human intelligence will arrive less than two decades (R. Kurzweil 2014)

Singularity has become an unexpected cultural reference. Ron Rosenbaum depicts a cadre of technologists who believe these ever-increasing AI capabilities will result in machine consciousness, and ultimately, that we will have the capacity to upload digital versions of ourselves to achieve immortality (Rosenbaum 2013). Kurzweil was interviewed by Rolling Stone's David Kushner in 2009 where he discussed an ambitious and personal plan to use the power of Singularity to reunite with his father, who died in 1970 (Kushner 2009). Rosenbaum describes these believers as techno-utopians. They have created their own form of spirituality, creating and worshiping Singularity, referring to their immortal dreams as the "Nerd Rapture" (Rosenbaum 2013).

Turing felt that the development of machines with a capacity equal to what Kurzweil calls Singularity was unlikely, if for no other reason than the societal consequences of building them. He imagined pushback from intolerant religious communities and "intellectuals who were afraid of being put out of a job" (A. M. Turing 1996). He aimed his sharp focus on what he felt were the ineffectual shortcomings of academics, writing that despite their effort to keep their standards of intelligence equal to a computer, it would not take long for an intelligent machine to "outstrip our feeble powers" (ibid. 1996). He imagined a world where the machines were immortal, conversing with each other to sharpen their wit, and ultimately taking control.

Reports describing recursive self-improvement, Computers As Social Actors, and the potential Singularity of AGI, have contributed to a flurry of articles with dystopian prognostications of a near-future run by sentient machines. The personal anxiety of living in a society dominated by ungoverned computers is a topic that can be effortlessly woven into fables of all types. We have seen representations in literature, film and gaming -- from the *Terminator*'s all controlling Skynet (1984) to the benevolent Thunderhead in Neal Shusterman's *Arc of the Scythe* series (2016). There are hundreds of examples where popular culture presents a technological future that appears to be around the corner.

While we *are* seeing a rapid acceleration in the capabilities of Artificial Intelligence, the development of AGI still sits on a distant time horizon. The required technology has yet to be developed. More importantly, to effectively model human intelligence we must first have a complete understanding of our inner workings. This includes an effective comprehension of intelligence and a definitive understanding of awareness. This will also require an understanding of the human mind's interdependence on the body's sensorimotor loop. Murray Shanahan describes this symbiotic mind-body relationship as an embodiment of human intelligence (Shanahan 2015). A human's mind is part of a body that has many dynamic parts. We have evolved to maintain our total well-being in an effort to perpetuate the species. This survival is predicated on situational awareness. Our bodies have muscles and senses that afford locomotion through the world, guided by an array of inputs. Shanahan writes that the brain sits in the middle of this sensorimotor loop, shaping our actions through perceptional feedback. Human intelligence, and the capacities for language, reason, and creativity all rest on this sensorimotor foundation. Currently, no aspect of Artificial Intelligence has the capacity to emulate the complete sensorimotor loop. The complexities of building a feedback network to sense, perceive, process, and learn from all stimuli, not just data but perceptual and environmental, far exceed the capabilities of our models.

Writer Bret Devereaux believes our development of AI may be more illusionary than actual, describing many of the announced advances in the modeling of human intelligence as ‘fake it till you make it’ (Devereaux 2023). He describes a theory where researchers intend to emulate intelligence, even though we lack the understanding of how the mind works in its full complexity. As a short-term solution, researchers have chosen instead to create machines which can convincingly fake being a mind in the hope that a “maximally convincing fake will turn out to be a mind of some sort” (ibid. 2023). He agrees that as a short-term solution this approach might work but posits that there is an equal likelihood it will not. He asks, what happens if this approach to machine learning is a dead end? Devereaux argues that researchers are presenting the illusion that we are far closer to creating a synthetic mind than is supported by fact. For AGI to be more than an algorithm on a machine it must work like human intelligence with the capacity to be aware. For now, human intellect is an enigma and a realization of AGI in the near future is still a fantasy.

The jagged trajectory of iteration in Artificial Intelligence's eight-decade history has not resolved the question of whether the human mind is truly computable. The strides towards emulating human intelligence are impressive, but the closer we come, the more difficult the solution. The inscrutability of AI has not, however, limited the deployment of working models. Compelling synthetic images, complex narratives, and original music can be created with existing algorithms. To generate a video story, a model will need to utilize the same computational logic found in applications like Stable Diffusion, GPT, text-to-speech, and DALL·E. This technology, Generative Artificial Intelligence, offers a promise that your computer will learn how to tell your story.

Chapter 4

Generative Artificial Intelligence and Video Applications

The developmental wave of Generative Artificial Intelligence is beginning to roll through the creative community. Synthetic sound, image, and video creation have the potential to disrupt representations of creativity, personal expression, and the ways we reflect our cultural values to ourselves. The technology promises to free us from the mundane so we can focus on the complex. It also portends a time when we devalue current modes of creative expression. How do these generative models work, how are they applied in video production, and what are some of the potential consequences of adoption?

4.1 Generative AI – Creative shortcomings and illusions of merit

Author Kevin Roose describes Generative AI as a “wonky” umbrella term (Roose 2022). It is a classification of algorithms which not only analyze existing data but apply what they have learned to create new media. Despite the constant flux of AI development and iteration, Generative Artificial Intelligence applications are finding their way to market. This includes many of the AI models we have previously touched on, including ChatGPT, DALL·E, Stable Diffusion, and CoPilot. Figure 26 is a snapshot from the end of 2022 illustrating how specific builds of AI were inserted into individual market sectors, and their applied purpose. The graph illuminates how text generation tools enjoyed both the quickest and widest deployment. It also points to the emergence of computer software code generators, image generators and speech synthesis. These examples demonstrate that as the platforms stabilize

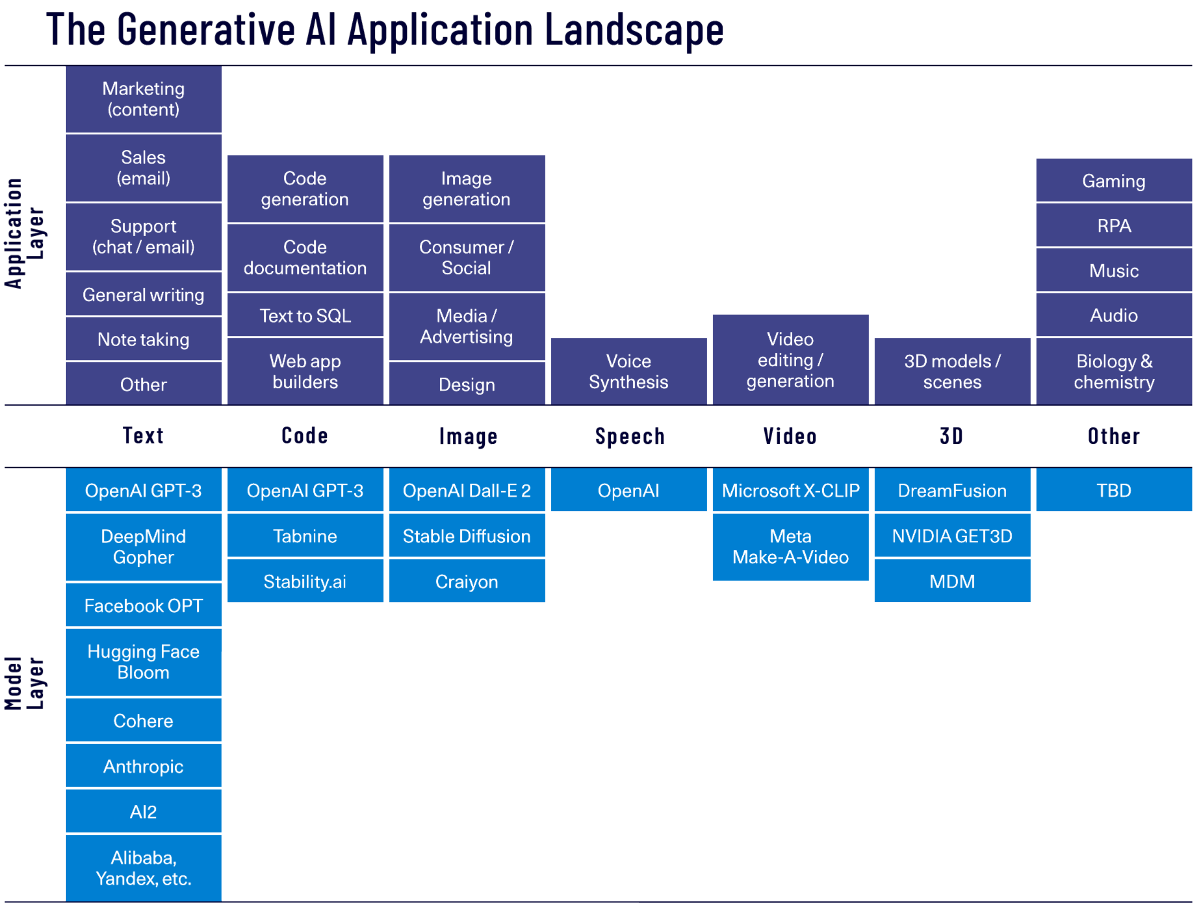


Figure 26 The Generative AI Application landscape
Adapted from: Generative AI: A Creative New World (Huang and Grady 2022)

developers have the room to focus on iterative quality improvements, emphasizing better, faster, and cheaper releases. Access to many of these models include free or open-source gateways with developer's expecting that through early adoption, users will demonstrate an explosion of creativity, ultimately growing the market (Huang and Grady 2022).

Perhaps the most widely talked about implementation of Generative AI, and the models with the widest market deployment, are text generators. These applications are examples of Large Language Models, an algorithm classification touched on in the previous chapter. In simplified terms, they process huge amounts of text in search of patterns, with the goal of applying the knowledge gained to generate statistically probable outputs. During their deep learning phase, they map and store instances of word placement. This would include how often specific words occur, the frequency of their conjunction with other specific words, and their relative positions in a sentence. It is important to note that the mechanics of this training do not include processing the context of the specific word, its meaning, nor the sourced author's thinking. It is simply the mathematical probability of a relationship between words.

For example, during training the model learns when it is presented with the phrase, “The brown fox”, there will be a very high degree of probability that the next word in the string should be “jumps”. The text output from these generators presents a humanlike presence through the appearance of intention and thought. But that appearance is an illusion.

Bret Devereaux notes these models accurately absorb their training text, yet they lack the capacity to understand or verify the source (Devereaux 2023). Additionally, they do not have the capacity to learn the definition, meaning, or etymology of words, nor do they understand the associations between the words used and their corresponding referents. Devereaux points out that the absence of referents is a critical point of failure in these models. Referents are the actual thing that a word denotes. We see letters on a page forming the word “cow” and we understand it is referring to a bovine animal that moos. We are likely to recall a mental association between the word and a personal experience. It might prompt a memory of a trip to the farm, a glass of milk or a stuffed animal you held as a child. This is a referent. These referents become stronger for us through our personal journeys acquiring a language. Our relationship with language is formed by the connectiveness of these referents. Devereaux believes the critical weakness of these Large Language Models is they rely on a matrix of statistical relationships that have been “stripped entirely of their referents” (ibid. 2023). Generated texts appear to carry meanings of intention and thought, but the veneer of intellect is only illusionary, stemming from the model’s fundamental inability to process meaning.

When we prompt a Large Language Model to write an essay there is implied that it forms the response by considering our request, researching the topic, gathering the evidence, to then craft an appropriate answer. The results from this prompt-and-respond cycle can feel authoritative to a user. This perception that a chatbot is a thoughtful, all-seeing, all-knowing font of information, a source of thought-through responses, is an illusion. Behind the curtain these algorithms sift through a mountain of words, choosing what it believes is appropriate based on statistical probability, to assemble a learned word order. In an essay for the *New York Times*, Noam Chomsky, Ian Roberts, and Jeffrey Watumull, write that the study of linguistics and the philosophy of knowledge teach us that these AI models differ profoundly from how humans use language and reason (Chomsky, Roberts and Watumull 2023). They note that the human mind is a surprisingly efficient system operating with only small amounts of information. In communicating ideas, we are not attempting to “infer brute correlations among data points, but to create explanations” (ibid. 2023). Devereaux describes the output of LLMs as little more than reproductions of training materials, lacking in original thought and an understanding of the underlying objects or ideas (Devereaux 2023).

It is clear the output from Large Language Models is little more than an assemblage of words, not ideas. This stems from the computer's paucity of understanding about the referents they use. Their capability is limited to imitation, lacking original thought, the assertion of truth, or logical argument. Humans use language to communicate, we write to convey ideas, we choose specific words to convey emotion, we craft word order to reflect the rhythm of our voice. The shortcomings of these models shortchange the experience of crafting a sentence or presenting original idea. As Devereaux observes, "It doesn't know anything about anything" (Devereaux 2023).

Generative AI is being applied in many instances beyond text generators. One example of an entirely different Generative AI model is the creation of images and video (Thompson 2022). Many companies large and small see significant potential in these algorithms, with millions of dollars poured into text-to-image generators in 2022 (ibid. 2022). We will discuss the mechanics and implications of these applications in greater detail later in this chapter, but it is relevant to introduce an emerging connection of LLM chatbots with Generative AI synthetic avatars. Applications like Chat D-ID afford a user the opportunity to talk directly with an AI chatbot (D-ID 2023). When asked a question, the application delivers the response through an AI generated on-screen character, talking directly to the user while the body of their text unfurls on the screen (*see figure 27*). Users can enter prompts via the text window or with their voice. When using voice prompts, the application leverages a speech-to-text engine that converts the user's voice into text. Chat D-ID is a coherent implementation of three AI models, Large Language, video synthesis and speech-to-text/text-to-speech. For a user, this integration of three discrete models delivers what can be described as a step towards a humanization of the AI experience. The opportunity to converse face-to-face with the avatar creates a more natural exchange between the computer and user, with the prompt and response flowing more like a natural

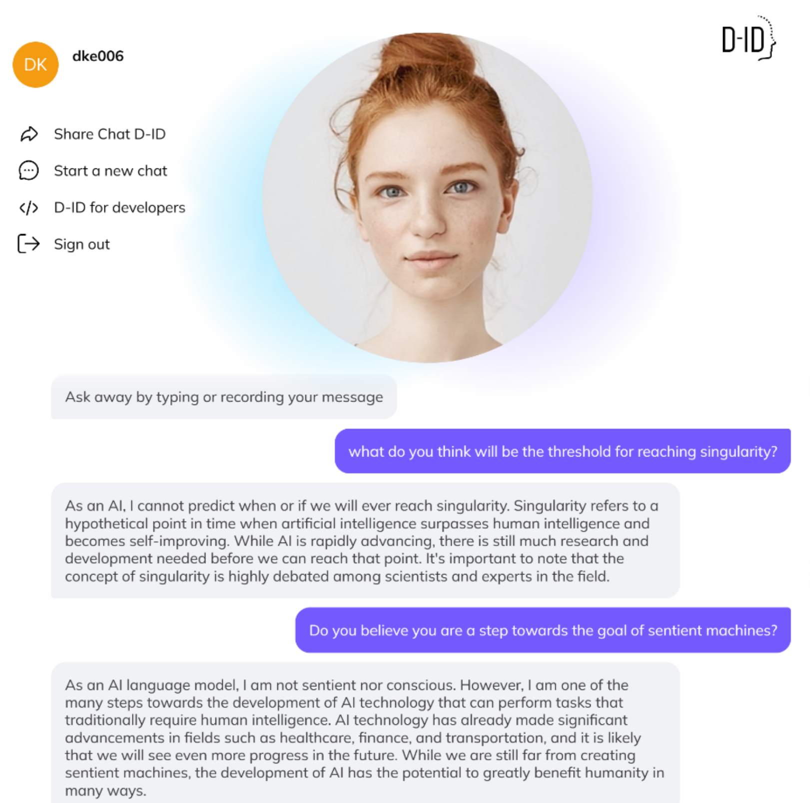


Figure 27 Chat exchange with AI generated Avatar on the D-ID platform. Captured March 9, 2023 (D-ID 2023)

conversation. One can easily fall into a conversation as if this were a person, unconsciously applying social rules and normative social behaviors. This blending of algorithmic logic with an illusion of humanity makes it difficult to avoid treating the avatar as anything but a human social actor.

4.2 GAN and CAN - Generative Adversarial Networks & Creative Adversarial Networks

Humans have demonstrated intelligence for centuries by creating visual images. As a species we convey ideas through the practice of pictorial style and technique, applied skills that give voice to an individual artist. The essence of painting is understanding the spectrum of artistic styles and possessing the knowledge to express one's creativity through a specific technical approach. AI image generators are attempting to replicate the human process by first understanding, and then applying, visual art's normative styles. To disentangle how these AI models replicate images it is important to first understand the nuanced difference between the two most widely leveraged algorithms, Generative Adversarial Networks (GAN) and Creative Adversarial Networks (CAN).

For nearly 10 years Generative Adversarial Networks have been the standard approach for generating images with Artificial Intelligence (Pound 2022). First proposed by Ian Goodfellow (Goodfellow, et al. 2014), the GAN is a method for predicting generative models using adversarial pathways (Goodfellow, Bengio and Courville 2016). Most of the previous AI models discussed in this paper, like Large Language Models, are classified as discriminative algorithms. They are trained on the specific features of a source dataset utilizing a linear path for classifying input data. This deep learning enables an algorithm to apply its knowledge for predicting a label or category when classifying other data sets. This process of classifying affords the model the ability to assign a numeric probability value indicating how, for example, a word would be used in a sentence, or which is the best next specific word in a phrase, or even if the most recent email in your inbox is spam. Discriminative algorithms are utilized to map the correlation between features and labels, while Generative algorithms take a different approach. Their task is not the determination of probability when using features to classify a specific label, rather they attempt to predict the features that are required to generate a label. Discriminative algorithms are about *finding* patterns, generative are about *creating* patterns.

The adversarial element of a GAN involves pitting two distinct models against each other and comparing their output. Comprised of two main components, a "Generator" network and a "Discriminator" network (*see figure 28*), the task of the GAN's Generator is the production of new data instances, while the Discriminator compares the Generator's output against images of the real item. If, for example, the goal is to generate an image of a mouse, the job of the Discriminator is to recognize

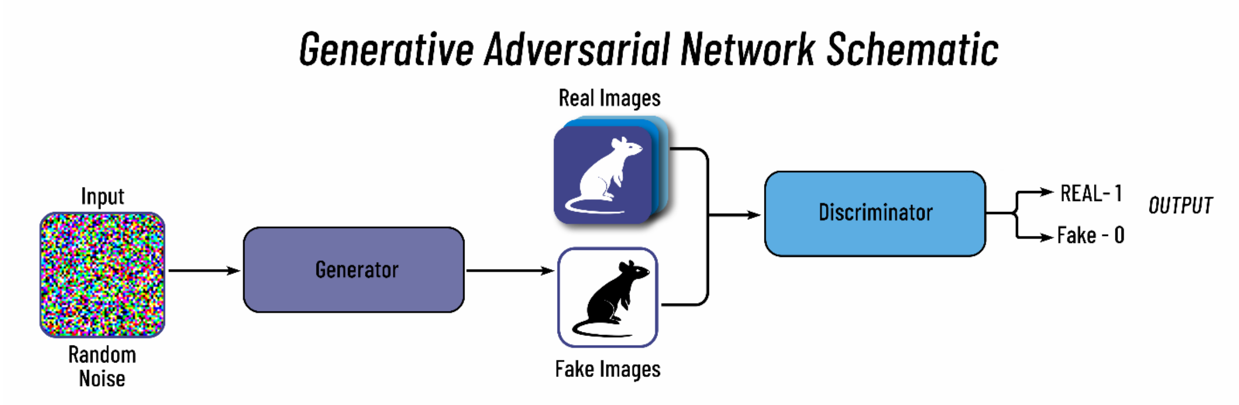


Figure 28 Generative Adversarial Network (GAN) schematic

authentic images sourced from the “real images” dataset, rejecting those it deems as fake. In parallel, the Generator is creating synthetic images of what it thinks represents a mouse, with these images also passed to the Discriminator for classification. The Generator’s task is to learn from the synthetic images that have been rejected and apply that knowledge for improved generation, ultimately producing something that is deemed authentic by the Discriminator. This adversarial game of cat and mouse is played with one network looking to generate fake images, and the other looking to separate real from synthetic.

The process begins with an image of random noise fed into the Generator as a seed. The generator processes the noise to create an image it believes matches the criteria of the requested classification. This image is fed into the Discriminator, along with a stream of images from the Real Images dataset. Using pattern recognition the Discriminator evaluates each image, returning a probability value between 0 (fake) and 1 (authentic.) This process cycles repeatedly, with the Discriminator operating in a feedback loop with the Real Images database, and the Generator in a feedback loop with the Discriminator, refining the generated image through the iterative application of the Discriminator’s feedback. This dynamic system has been described by David Pfau and Oriol Vinyals as an actor-critic model, where the Generator is learning how to effectively trick the Discriminator, and the Discriminator uses feedback to constantly refine the filters used for classification (Pfau and Vinyals 2016). Through competition each model improves, with the synthetic instances of generated data ultimately imitating the original source dataset. The goal in this example is to generate an image of a mouse that convincingly looks like a mouse.

The process of training a GAN is inherently difficult. One problem is an issue referred to as mode collapse, where the system repeatedly generates the same image, with no improvement or variety. These algorithms are trained to solve a specific problem: generate an image defined by a prompt. There are no inherent incentives for the GAN to generate refinements once it has created something that the Discriminator judges to be “real” instead of “fake”. As Google Scholar and University of Nottingham professor Michael Pound points out, “Once it finds the correct solution to the problem it simply repeats it (Pound 2022).

These models are not always successful at imitating life. As a point of fact, releases like DALL·E 2 and Stable Diffusion v2.1 shared common problems when attempting to render lifelike human forms (*see figures 29 and 30*). AI generated images are based on a stochastic approach where the algorithm learns from random elements being inserted during the generation process. All stochastic practices involve a large number of inputs or trials, with an expectation that the model will attempt a wide array of approaches as it searches for the best solution. By increasing the number of trials, you increase the accuracy of the results. For generating an image, it leverages what it has learned from the training dataset, beginning with primitive elements, and refines them probabilistically with iterative cycles. Because data-centric stochastic machine learning is a model that relies on randomness it is unable to guarantee accurate output. An inherent problem with this random trial-and-error approach is during the refining stage it may choose a path it believes is correct, but once completed, the output reveals

Figure 29 Rendering of misshapen human features using Generative AI
Images generated: Lensa, December 26, 2022 & DALL·E 2, January 20, 2023

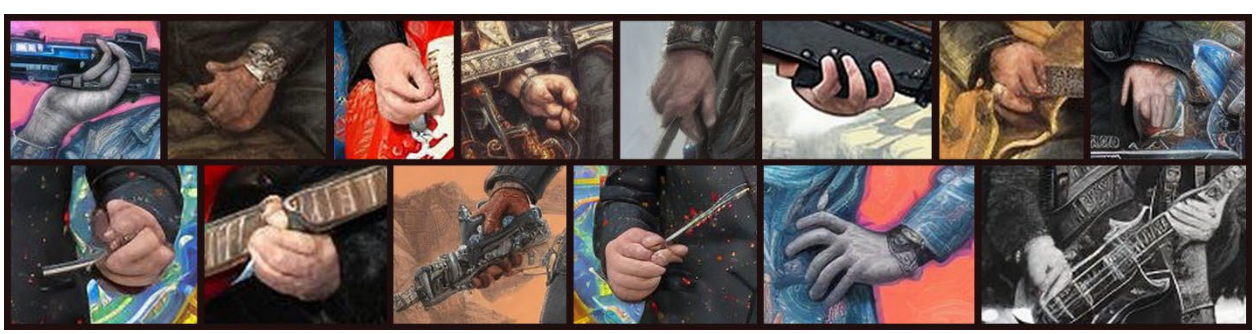


Figure 30 Lensa rendering of misshapen hands and fingers using a Stable Diffusion model.
Images generated: December 26,2023

mangled and misaligned human forms. The refining inaccuracy can result in a-symmetrical eyes, circles drawn as ovals, and rather comical renderings of human faces and hands.

Like most machine learning software, these models work by identifying and replicating patterns in data. To train a GAN you must first gather millions of images. These are used as the Discriminator's source dataset during deep learning, informing the algorithm as it connects specific images with text-based keywords (Pound 2022). In Stable Diffusion's case, this popular image generator was trained with three massive datasets collected and distributed by the Large-scale Artificial Intelligence Open Network (LAION). LAION is a worldwide nonprofit organization with the stated goal of "Making large-scale machine learning available to the general public" (LAION 2023). The images comprising these collections have been scraped from the internet, with most sources created by human artists and photographers, and a significant percentage of them protected by copyright (Vincent 2022). Gathering and organizing billions of images requires significant computing resources on very large networks. The costs association with this curation can be significant. During the process of gathering and organizing these datasets, LAION's compute time was largely funded by the company that would use it for training, Stable Diffusion's owner, Stability AI (Baio 2022).

Each of LAION's image datasets are originally compiled by Common Crawl, a nonprofit that has established vast repositories of web files. The organization scrapes billions of webpages monthly, releasing the results as massive datasets available for download from their website (Common Crawl 2023). To create a collection of images that would be applicable for training an application like Stable Diffusion, LAION first sifted the data from Common Crawl for HTML image tags, searching for images that had alt-text attributes. To filter the resulting 5 billion image-pairs into separate datasets, they then classified each image by technical metrics, including language. The outcome of this filtering resulted in a variety of collections, each with image totals numbering in the tens-of-millions. These collections included classifications sorted by image resolution, the likelihood a watermark, and each image's subjective visual "aesthetic" score (Baio 2022).

Initially Stable diffusion was trained on a set of low-resolution 256-pixel-by-256-pixel sized images, 2.3 billion English-captioned images, and high-resolution images that had been culled from a larger collection of 5.85 billion image-text pairs. By the summer of 2022 the Stable Diffusion model had been repeatedly refined via additional training, with the most recent round of training leveraging a set of 600 million high quality aesthetic images (ibid. 2022).

Researcher and technology writer Andy Baio wanted to explore if these images were being sourced from a broadly dispersed sample of random websites. He was uncertain if there were individual internet image repositories that were experiencing a disproportionate number of their images being downloaded and used for training. For creators, this question is of particular interest because many websites include images shielded by copyright protection. Baio initially indexed 12 million images, filtering by their domains and noting the source website that originally hosted the file. While there were 12 million images in the dataset, nearly 47%, of them were sourced from only 100 domains. This high percentage reveals that social media, shopping, and photo/art websites were disproportionally represented (*see figure 31.)* Baio notes that it was not entirely surprising that the social media site Pinterest, and the blogging platform WordPress topped the list. What was significant to him is that nearly six percent of the 12 million images in the set were sourced from Fine Art America, an art print and poster website (ibid. 2022). Shopping sites like Shopify, Wix, Squarespace and Redbubble also totaled six percent of the images. Additionally, Stock photo and artwork sites, including 123RF, Adobe, PhotoShelter, iStock Photo, and others, were disproportionately represented.

Questions about the source of images used in training Generative Adversarial Networks are not limited to Stability AI's Stable Diffusion. In the paper, "Hierarchical Text-Conditional Image Generation with CLIP Latents", the research team at OpenAI writes that their image generator DALL·E 2 was trained

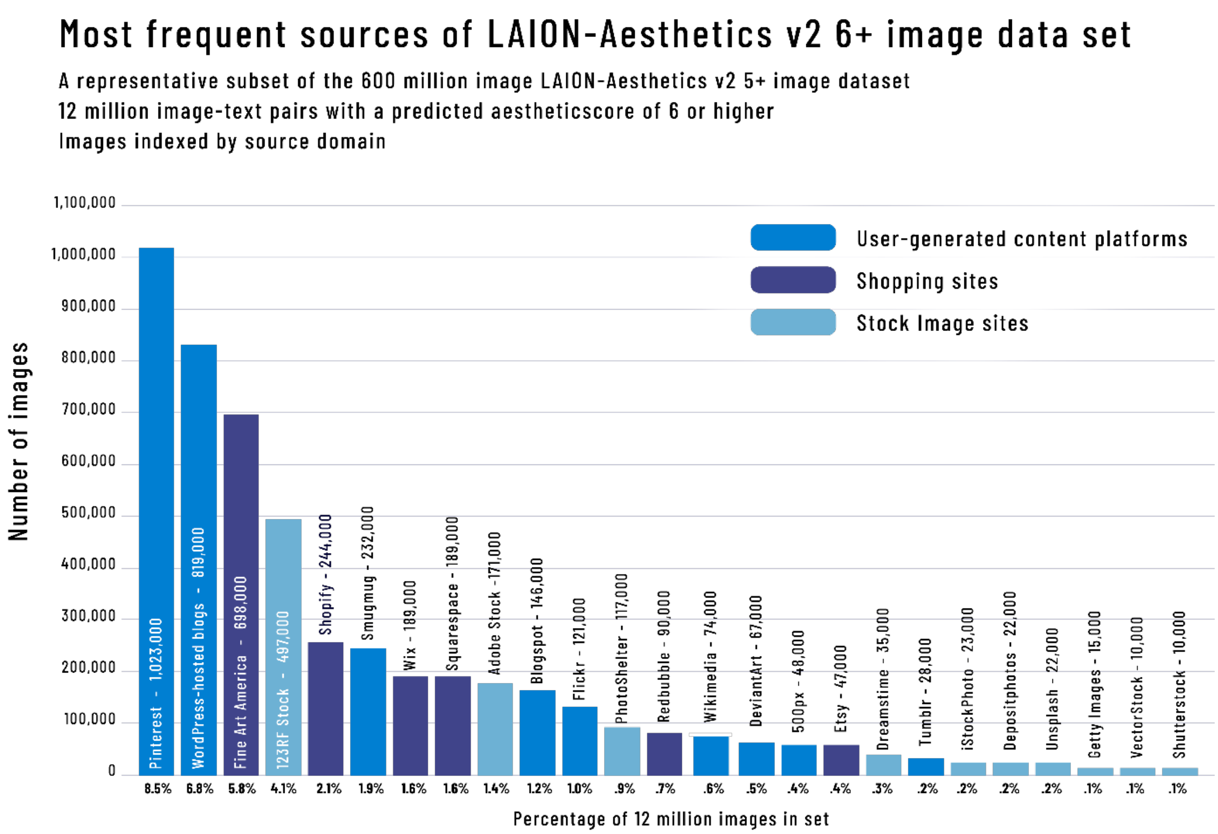


Figure 31 Source data: Andy Baio - waxy.org 30 August, 2022

on approximately 650 million image-text pairs also scraped from the Internet (Ramesh, et al. 2022). They leveraged an alternative image dataset to LAION, ImageNet, which is coordinated by researchers at the American universities Princeton, Stanford, and the University of North Carolina in Chapple Hill (ImageNet 2021). Initial ImageNet models were trained on datasets that were manually labeled, an expensive process that required over 25,000 workers who annotated 14,197,122 images for 22,000 object categories (Radford, et al. 2021). Like Stable Diffusion's training, OpenAI's CLIP algorithm, running on a GPT-2 model, processed subsets of the data to learn the relationships between images and the words used to describe them (Strickland 2022).

The utilization of copyrighted images and artwork in training AI models raises concerns about potential legal infringement on the creator's rights (Wiggers 2022). The training data utilized by companies such as Stability AI and OpenAI may have violated Intellectual Property law by replicating protected art, photos, graphics, and images, leading to theoretical problems in several aspects. In an article written by TechCrunch reporter Kyle Wiggers, Intellectual Property attorney Bradley J. Hulbert observes artwork that is "demonstrably derived" from a "protected work" has generally been found by the courts to be infringing, even if additional elements were added (ibid. 2022). To be shielded from copyright claims, the work must be transformed to such a degree that the IP is unrecognizable.

AI image generation excels at creating unique images of specific objects rendered in a particular style. Most of the styles requested by users are not universal classifications like "abstract", "impressionist", or "Baroque". Rather, they are the personal styles expressed by individual creators. Users request images to be created in the style of specific artists such as Banksy, Diane Arbus, and Thomas Kinkade. On a growing number of repositories, like "Midjourney and DALL·E Artist Styles Dump with Examples: 130 Famous AI Painting Techniques*",* you can find list recommendations for which artist to request when crafting a prompt for an image. These repositories provide tables with lists of living and deceased artists ranked by popularity, sorted by style categories (Yalalov 2022). From styles like Figurativism to Retrowave to Crayon Art, living artists like Japanese painter Yoshitaka Amano to graphic artist Karol Bak to digital artist Mike Winkelmann (Beeple), these tables offer users a streamlined path for creating media that is strikingly similar to the original artist's work. Legal experts fear that the media created by these diffusion image models may lead to either a devaluation of the artist's work or confusion over what is genuine artwork, as the media produced is comparable to copyrighted work (Agarwal and Varshney 2019). AI-generated artwork that infringes on an independent artist's work could threaten their livelihood (Wiggers 2022). Therefore, the implications of training a system to replicate protected art, photos, graphics, and images can be considered a potential legal infringement on the creator's rights.

It is worth noting that the images used to train these Artificial Intelligence systems often display their provenance in the generated output. This observation brings to light that many of the source datasets used to train these systems contain digital representations of copyrighted artwork, complete with artist signatures. As is customary in the practice of painting, human artists often study the works of others to learn and refine their style and technique, drawing inspiration from the artistic influences they encounter. The question arises then, where is the boundary between an AI model being influenced by a specific artist as opposed to specifically copying their work (Pueringer 2022). Given that diffusion models are trained to replicate patterns within a particular style, they not only reproduce elements such as color, lighting, compositional framing, backgrounds, and line work, but also attempt to include a signature (Brownlee 2022). Figure 30 illustrates that the Prisma application, Lensa, considers signatures an important element to include when creating personal portraits, implying that the system was trained on original work created and owned by artists. When Lensa appeared in late 2022 and became the number one app in mobile device app stores (Ceci 2023), artists expressed concern that it offered little protection from their artistic voice being sold to a mass market without compensation (Harrison 2022). For a nominal fee, users could upload ten selfies, and their personal photos would be used to train the model. Hours later, the company would deliver 100 AI-generated personal avatars in various settings. Many of the uniquely generated portraits include graphic elements that appear to replicate an artist's signature (*see figure 32*.) Molly Crabapple describes the output from these applications as troubling, because "AI mashes together art painstakingly created over lifetimes, then spits out an image, even mimicking an artist's signature" (Crabapple 2022). The Lensa app serves as an example that it is very likely artists' work is being used without their consent to generate income for a third party.

By now you are familiar with the parts and process of various AI models, particularly the Generator and Differentiator networks that form the heart of a GAN. Like other models, Generative Adversarial networks suffer from computational limitations. Over time the Generator acquires a level of

Figure 32 Signatures included in portraits generated by a Stable Diffusion model. Images generated: December 26,2023

competency where it becomes very efficient at quickly creating images that fool the Discriminator. While it is successful at generating new images it follows a computational path of least resistance because it is narrowly trained to solve a specific problem. The results are appealing yet derivative, and rarely novel or innovative. Demonstrating human levels of creativity requires an entirely different construct (Cheng 2022).

When looking at art we are not looking to merely identify a flower or a dog or a person. The experience is much deeper, with nuanced layers of understanding. Artwork can elicit strong emotional responses. It can provoke, inspire, or wash waves of tranquility over a viewer. Yet much of the work generated by AI lacks the depth and complexity of human created artwork. Most work can be considered little more than a shallow representation of a thing, rather than expressing an emotional point of view. To be a creator, to create expressive art, AI must first understand the act of creation through differentiation and then execution. Before these models can prove intelligence equal or surpassing us, Ahmed Elgammal believes AI must have the capacity to demonstrate human-level creativity (Elgammal 2019).

Elgammal is a computer science professor at Rutgers University, and director of their Art and Artificial Intelligence Lab. To create a model that could potentially demonstrate human levels of creativity, in their paper, "CAN: Creative Adversarial Networks, Generating 'Art' by Learning About Styles and Deviating from Style Norms", Elgammal and a cohort of students introduced a novel concept (Elgammal, Liu, et al. 2017). Elgammal and his team proposed a system that would not only look at millions of individual pieces of art to learn artistic styles, but they included a mechanism that asked the model to slightly deviate from a particular style, innovating new art through its execution of artistic technique. The goal is to create art that, because of its novelty, prompts a higher level of interest and engagement in a viewer (ibid. 2017). They approached this problem by linking the algorithm's creative process through a model that mimics our species' development of art over time. Their objective was to simulate exactly the process of humans developing art.

Throughout their lives, artists are exposed to various forms of art, forming an impression that informs their creative efforts. This results in work that not only expresses their personal viewpoint but affords them the capacity to evolve their artistic perspective and technique over time. The team at Rutgers proposed a modification to the GAN's objective, enabling it to generate creative art by maximizing the deviation from established artistic styles while minimizing deviation from the original art used in the Discriminator phase (ibid. 2017). Like a GAN, these models operate via deep learning on neural networks, but unlike a GAN the output of a Creative Adversarial Network is not specifically imitating artwork that currently exists.

In his article, “Biological Bases of Creativity”, Colin Martindale argues that humans find art attractive when they are aroused by something novel (Martindale 1999). This push against habituation increases interest when something is perceived as exclusive, yet within a historic context. A CAN is looking to satisfy a viewer’s “arousing potential” when viewing the generated image. This potential rise and fall of interest can best be illustrated with a Wundt curve.

In 1874 Wilhelm Wundt introduced the concept of “optimal level of stimulation” (Wundt 1874). He plotted a modified bell curve relationship between Stimulus Intensity and Hedonic Value (*see figure 33*). His curve illustrates how many forms of stimulation are pleasant at lower, controlled intensities. As the stimulation increases the feeling of pleasantness increases until it reaches a point where the intensity is too high, prompting an unpleasant response. In the 1970s Daniel Berlyne built on Wundt’s work, forming his theory of “intermediate arousal potential” (Berlyne 1970). Berlyne was investigating how viewers would rate colored shapes on scales of “pleasingness” and “interestingness”, corelating the results against their ratings of “novelty.” His results demonstrated that, on the whole, “both pleasingness and interestingness increase with novelty” (ibid. 1970, p. 284). The push for novelty cannot be too extreme, though. If the level of arousal is not controlled, continuing to grow exponentially, a viewer is likely to form a negative reaction. One’s sense of pleasantness is lowest when we are relaxed or sleeping. When an event or object prompts low levels of arousal it can be perceived as boring. It reaches maximum levels when we confront situations that are perceived as dangerous, violent, or confrontational. When the arousal is extreme it can prompt a hostile or fight-or-flight response. Berlyne found that only intermediate stimulation results in curiosity. The extreme ends of the Wundt curve reflect the two rules in stimulus selection: Avoidance of Boredom (AoB) and Avoidance of Anxiety (AoA) (Wu and Miao 2013). A CAN is looking to satisfy a viewer’s

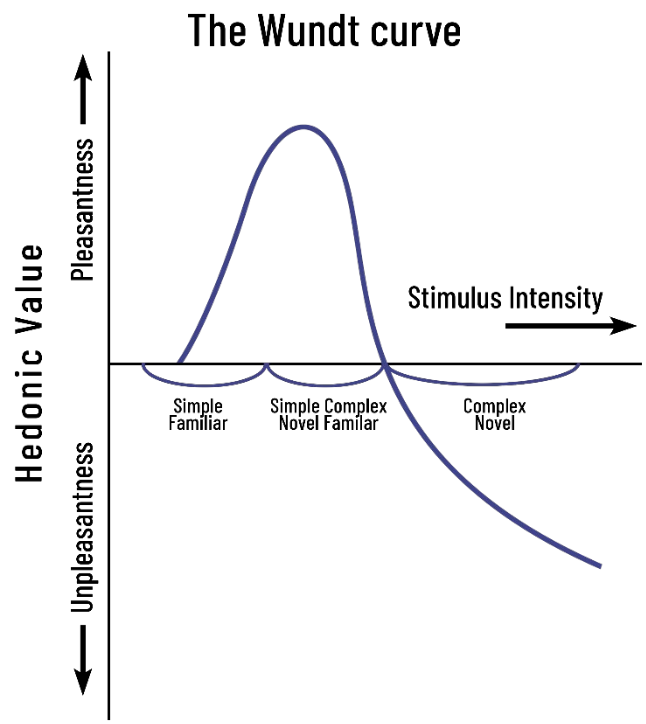


Figure 33 The Wundt-curve according to Berlyne (Adapted from Wundt, 1874, p. 432 and Wu and Miao, 2013, p. 18:7).

“arousing potential” when viewing the generated image, while neither boring them nor pushing image aesthetics beyond their understanding or appreciation of a specific style of art. The hope is it will pique a viewer’s curiosity.

On a fundamental level, Creative Adversarial Networks share a common architecture with the GAN model. The task of a Discriminator is still to learn how to classify images as being genuine (real) or synthetic (fake). The training of a CAN is more complex because the algorithm learns not only by labeling specific objects, like a GAN, but they also learn how to classify each image according to its visual style. This process is a critical step in affording the algorithm the structure for forming an understanding of creativity through the development of concrete definitions, process, and structures in art. Similar to decision trees, they require this type of intellectual scaffolding to effectively mimic the creative process. The classification of the human-created art into styles includes categories assigned to 25 different specific schools and techniques (Elgammal, Liu, et al. 2017). For example, styles could include labels like Cubism, Impressionism, Abstract, Romanticism, High Renaissance, Realism, Pop Art, and Primitivism. Ultimately, this understanding of classification of styles should enable the generator to think creatively.

An important characteristic of a CAN is that it learns the history of art through the process of creating art (Thoutt 2017). It is continuously learning as it is introduced to new concepts and styles, adapting what it generates by applying the knowledge gained. While the model has the capacity to apply this new knowledge, a Creative Adversarial Network shares the same shortcoming of other AI models. It lacks any semantic understanding of human art (Elgammal, Liu, et al. 2017). It is operating without context, nor does it have an explicit understanding of the elements depicted in the art or the principals of composition, color, or form. At a fundamental level, a CAN has a comprehensive understanding of how to replicate the patterns and elements within a specific category but has no clue why one artist paints romantic landscapes and another expresses themselves through abstract impressionism. The social, cultural, and historical contexts that frame our understanding of creativity are absent.

Like a GAN, the core of a Creative Adversarial Network is made of two adversarial networks, a Generator and Discriminator (*see figure 34*.) The model is modified by feeding the Discriminator millions of images with specific style labels that define the category of artwork. Here the Discriminator not only learns to identify an example of a “real” object or subject, but how to label the various characteristics of specific styles.

Again, like the GAN model, the CAN’s Generator seeds the process with an input vector of random noise. To create its image the Generator does not have the same access to art that is leveraged by the Discriminator. Instead, it relies on a modified feedback loop with the Discriminator. Each time it

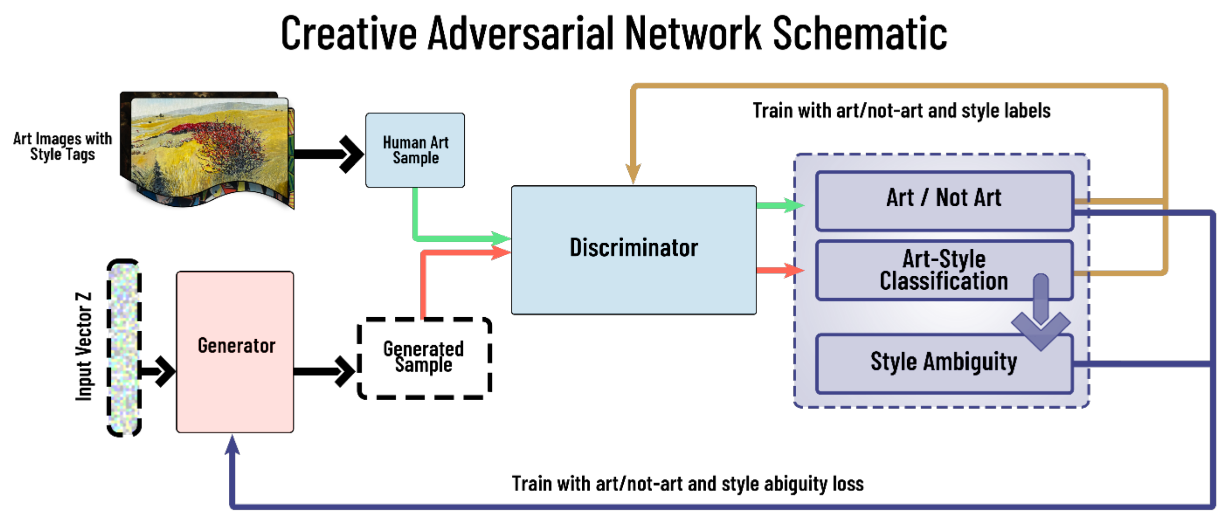


Figure 34 Creative Adversarial Network Schematic, Adapted from Elgammal, Liu, et al. 2017, p. 7

generates an image it receives two discrete signals from the Discriminator, the first specifies whether the output is “Art” or “Not Art.” This signal indicates if the generated image fits within a criterion that matches the Discriminator’s human art sample. In a traditional GAN system, this signal would be used by the Generator to adjust the weights to create an image that is more likely to deceive the Discriminator. Here the signal is prompting the Generator to converge what it is generating into the type of images it has already generated and that fooled the system, accelerating the process by applying what it has learned. The second signal informs the Generator how accurately it has created an image that falls within one of the 25 defined categories.

The Generator treats these two signals as opposing forces. It measures if the Discriminator has the capacity to measure the fidelity of the image and how it classifies it within a stylistic category. The effect of the second signal is it’s pushing the Generator to create work that is stylistically ambiguous, but when it strays too far the Discriminator penalizes it heavily. A simpler explanation is the first signal tells the Generator to color within the lines and the second tells it to break the rules. The Discriminator is rewarded when it forces the Generator to produce images that cannot be easily classified into a particular artistic style, with the output being images that are truly unique. The Generator’s ultimate goal is to create “fake” images that will be classified as “real”. The adversarial game of generate and test, of simulate and reject, of mixed signals, pushes the Generator to explore established creative spaces from the 15th to the 21st centuries while generating ambiguous and ambitious art.

A central motivation of a Creative Adversarial Network is the premise that viewers will perceive acts of creativity when a given piece of art is unique, without straying beyond the historically accepted bounds of a specific style. This AI model is attempting to increase stimulus intensity, thus increasing pleasantness, without straying into overly complex or uncomfortably novel artistic representations of what is often accepted as art. But the work is not simply pleasant, it is a unique expression within a style. The model's architecture affords it the ability to simulate what humans perceive as creativity. According to four human subject experiments conducted by Elgammal, participants who reviewed both AI and human-created art not only believed the AI generated artifacts were created by human artists, they also rated them higher on evaluative scales (Elgammal, Liu, et al. 2017). 85% of subjects in one test rated the CAN created artwork as being created by a human (ibid. 2017)

Elgammal believes for a system to establish that it is creative it should demonstrate three things: imagination by producing novel artifacts; skill by creating artifacts of quality; and it must possess the ability to assess the creative value of its output (ibid. 2017). He points to the adversarial communication within the CAN's algorithm, thus forcing it to explore creative spaces, as a demonstration of imagination. Furthermore, he believes it is imaginative since it is finding solutions that deviate from established styles while staying within aesthetic boundaries. He notes that the feedback loops provide a mechanism for self-assessment, improving and refining its output as it iterates images. Finally, he believes the model demonstrates artistic skill because human evaluators have determined the work generated by AI meets or exceeds the efforts of human artists. Their Creative Adversarial Network produced results where many of the evaluators not only thought these artifacts were created by human artists, but also rated them higher (ibid. 2017).

4.3 What is Diffusion? (DALL·E 2)

Competing AI architecture is not unique to the tug-of-war between symbolic and neural AI. With image generation there are alternatives to the adversarial network approach discussed above – one of the most successful models is Diffusion. Specifically, OpenAI's development of Contrastive Language–Image Pre-training, or CLIP. OpenAI chose to leverage diffusion because it affords them the opportunity to integrate it with their deep understanding of Natural Language Processing. Their image generation implementation of CLIP and diffusion, called DALL·E, runs on a 12-billion parameter version of GPT-3 (Ramesh, Pavlov, et al. 2021). In their paper, "Zero-Shot Text-to-Image Generation", Aditya Ramesh, Mikhail Pavlov, et al., present a transformer language model, trained to generate images from text descriptions. Like the Generative Adversarial Network, it utilizes a dataset of text-image pairs for deep learning. The system architecture incorporates techniques from machine learning, computer vision, natural language processing (NLP), natural language generation (NLP), and synthesis (ibid. 2021).

OpenAI recognized the inherent problems associated with compiling vision datasets, including the prohibitive costs in labor and time (Ramesh, et al. 2022). This approach of training on a specific corpus of data also suffers from the limitation of teaching a narrowly defined set of visual concepts. For example, Elgammal's Creative Adversarial Network is restricted to defining 25 categories of art styles. Teaching a GAN or CAN to visually read images results in models performing only one task well. Adapting a model to a new task requires significant effort through additional training. Additionally, these models often display performative problems, running well during benchmark testing but operating poorly in stress tests and real-world applications (Ramesh, Pavlov, et al. 2021b). For some, this litany of challenges has cast doubts on the prospect of using a traditional deep learning approach when building computer vision models (Radford, Sutskever, et al. 2021).

Because DALL·E is a transformer language model it does not require the same labor and computing time necessary to classify the millions of alt-text/image pairs. Rather than relying on 25,000 workers to classify 14 million images (Ramesh, et al. 2022), or the massive processing time required by LAION to organize their dataset (Baio 2022), OpenAI designed an efficient neural network where CLIP learns visual concepts via supervision by their GPT natural language model (*see figure 33*). The model is composed of a text encoder and an image encoder, which are trained on a large, heterogeneous collection of image-text pairs (Ramesh, Dhariwal, et al. 2022b). The encoders map their respective inputs to points on a "concept space", with this embedding information shared by both modalities. When processing data they map the image/text instances that match to points in the concept space that

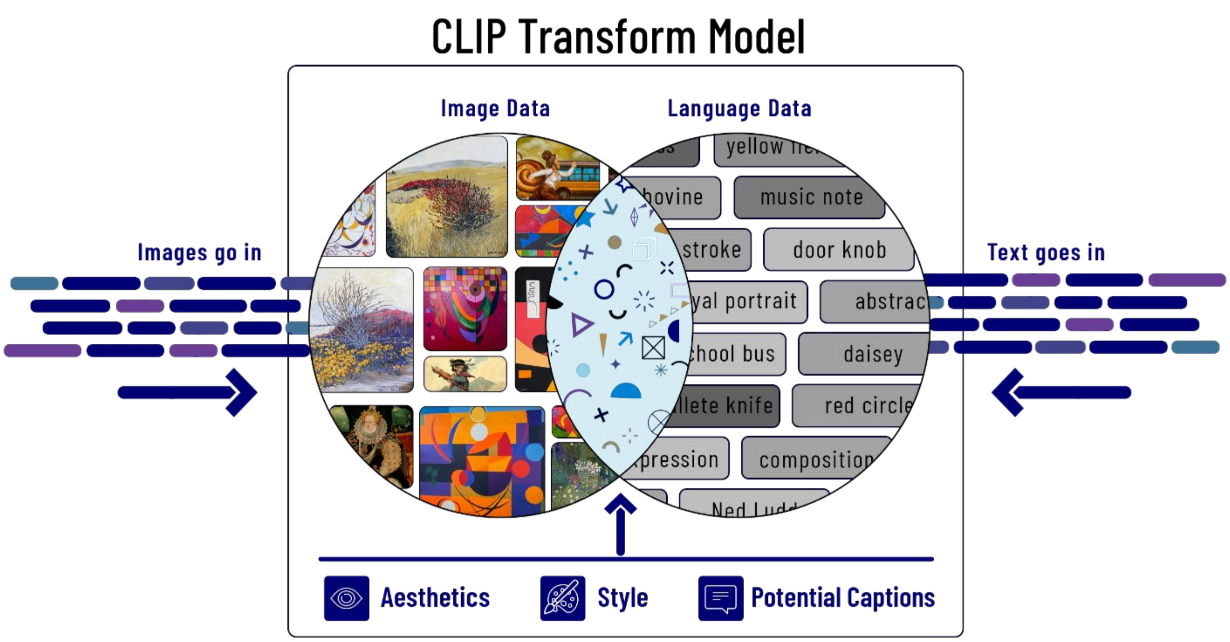


Figure 35: DALL·E Deep Learning via the CLIP Transform Model, Adapted from: "The REAL fight over AI art", 2022

are in computational close proximity. They assign mismatching pairs to remote points in the concept space. This contrastive training encourages CLIP to learn features about the images that people are likely to write about. DALL·E creator Aditya Ramesh described the process as, “If you imagine a Venn diagram where there's one circle for all of the information that's in images and another circle that holds all of the information that's in language, it learns the part in between” (Ramesh 2022).

Culling this intersection of image and language affords CLIP the ability to correlate relationships, acquiring computational associations of aesthetics, style and a broad knowledge of descriptive language. CLIP can be applied to any visual classification benchmark by providing the names of the visual categories to be recognized, leveraging the “zero-shot” capabilities of GPT-2 and GPT-3 (Radford, Sutskever, et al. 2021). The power of DALL·E 2’s CLIP architecture stems from the utilization of GPT as the scaffolding to process the massive quantity of paired natural language/image data scraped from the internet (Ramesh 2022). Automating the process of filtering and organizing this data not only removes the developmental chokepoint associated with manually labelling datasets that was described earlier; but employing an uncurated dataset containing the inherent noise and clutter of real-world sources for deep learning results in benchmark performance that better reflects the type of data these models will encounter. Developers believe this approach addresses the limitations of other generative models by reducing labor and expense, increasing task flexibility, and the demonstrable ability to perform equally well outside of controlled environments.

DALL·E is trained to generate images from text descriptions, receiving both text and image as a single stream of data containing up to 1,280 tokens. Aditya Ramesh, Mikhail Pavlov, et al. describe a token as any symbol from a discrete vocabulary (Ramesh, Pavlov, et al. 2021). As humans, we leverage individual letters in an alphabet as language tokens. DALL·E’s vocabulary can be considered more abstract, with tokens for not just word parts and text characters, but also image concepts. Of the 1,280 available tokens, a text/image pair would use a maximum of 256 for the text caption and 1,024 would be available for defining/describing the image concepts (ibid. 2021). Because DALL·E is a relatively simple decoder-only transformer, receiving both the text and the image as a single stream, it is trained to generate tokens sequentially. This autoregressive model allows DALL·E to both generate an image from scratch, and to regenerate any bottom-right corner rectangular region of an existing image (ibid. 2021).

In addition to benefits like cost effectiveness, efficiency and flexibility, Ramesh notes that CLIP models share the benefits of a Creative Adversarial Network, demonstrating the capability to understand specific image styles. Where the model differentiates itself from a CAN is the capacity to learn representations of image semantics. This understanding affords it the skill to vary non-essential details and generate multiple variations of an image based on the initial prompt. (ibid. 2021).

The structure of DALL·E is a hybrid model that combines three algorithms to perform the single task of image generation. It leverages GPT's Large Language Model for supervision, the CLIP model to learn and then apply that knowledge, and a Diffusion model, called unCLIP, for image generation. OpenAI believes this approach results in better photorealism and photo fidelity, while requiring a limited number of samples (ibid. 2021). The DALL·E architecture has several moving parts, with instructions and feedback transmitted between the networks. For insight into the process, and to disentangle this communication, it is important to first define the algorithmic process of diffusion.

Diffusion was initially inspired by research in thermodynamics. Early generative machine learning models, including both likelihood-based models and Generative Adversarial Networks, had inherent limitations when source data was cluttered because of low thresholds of information. To separate the data from the clutter, researchers "perturbed" the data with random and various magnitudes of Gaussian noise (Song and Ermon 2019). They passed this noisy data stream through a network that had been conditioned on the same levels of noise, filtering out the clutter that had been added, to generate an estimate of the original data's value. This method of denoising a Gaussian source can also be used as a process for generating images. Diffusion models are trained by gradually and systematically adding Gaussian noise to an image using a model that applies a parameterized Markov chain (Ho, Jain and Abbeel 2020). With the knowledge gained from the process of adding noise, the algorithm can be reversed. To produce an image from a source of pure Gaussian noise the model inverts what it has learned, reversing the steps to sequentially remove noise in a series of time steps (*see figure 36*).

DALL·E augments the Diffusion Model's training with additional textual information from CLIP, ultimately resulting in text-conditional image generation. Depending on training, a Diffusion Model can reliably generate photorealistic images of any object or environment. The process of image generation

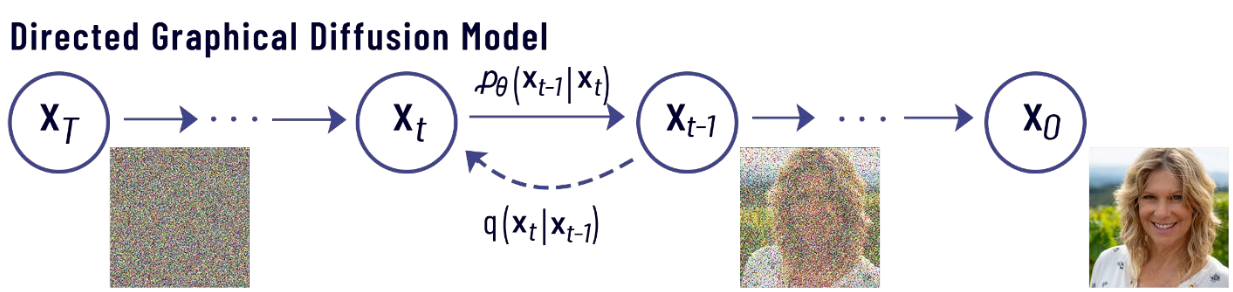


Figure 36
Adapted from: Denoising Diffusion Probabilistic Models, Ho, Jain, Abbeel, 2020

begins when a user describes a scene via a text prompt. Applying what it has learned from training, the algorithm generates a list of features that could be included in the image. These tokens are elements in the scene, like the point of a cat's ear, the texture of a woman's hair, or the edges of an apple. The tokens are fed into the diffusion model, generating the pixels needed for each feature. Using feedback from the Discriminator the algorithm iterates the individual pixels into a coherent image that is both semantically consistent and a photorealistic response to the user's prompt. In an article on his website, Aditya Ramesh describes the process as making the image incrementally realistic, "eventually yielding a pristine, noiseless image" (Ramesh, Dhariwal, et al. 2022b).

To improve the fidelity of image generation stochastic noise is inserted at random points during the phase of iterative image refinement. While this approach delivers better pictures, one consequence is DALL·E can suffer from some of the same problems seen in other diffusion models, with multiple fingers on human hands, misshapen faces, additional limbs, and asymmetrical eyes (*see figures 29 & 30*). A benefit of using the stochastic process is the model creates variations of the same image by inputting the same encoding vectors through the algorithm multiple times.

Once it has been trained, DALL·E operates as a two-stage neural network, first by generating the "gist" of an image through the CLIP model and then filling in the remaining details with the unCLIP Diffusion model, resulting in a realistic image (*see figure 37*) (ibid. 2022b). unCLIP uses the CLIP text prompt, image concept, stochastic variables, and Gaussian noise to generate four image variations (Abram 2022). Figure 38 illustrates both sides of this process in detail. CLIP's initial deep learning appears above the dotted line, where it is trained on image/text pairs. Here the text and image encoders use contrastive training to map the shared representation in the "concept space".

Image generation begins with a prompt from a user that seeds the process for generating a CLIP text "embedding". This embedding becomes the computational objective for the algorithm, a representation of CLIP's understanding of the prompt based on how weighted tokens have been mapped in the concept space. The embedding is sent to unCLIP for image generation and refinement.

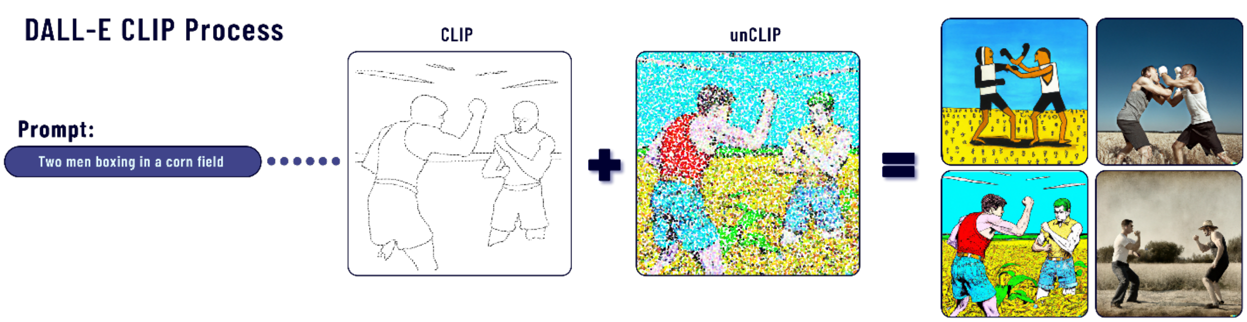


Figure 37 DALL·E CLIP- unCLIP process

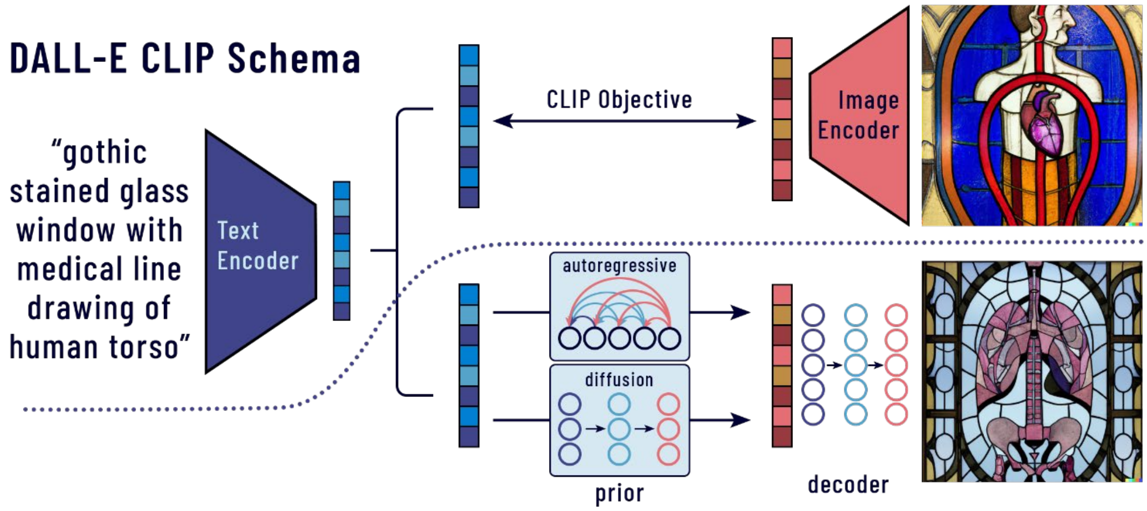


Figure 38: DALL·E CLIP Schema Adapted from: Ramesh, Dhariwal, et al., Hierarchical Text-Conditional Image Generation with CLIP Latents 2022, p. 3

Where the CLIP model outlined the framework of the image, the gist if you will, unCLIP fleshes out the details. It is important to note that the CLIP model remains static during the next two phases of the process. When the tokens arrive, they first go to the prior model as instructions for training. The tokens cycle through the prior's autoregressive engine and diffusion model to generate a map of the corresponding image encoding. With the reverse diffusion steps defined by the prior, DALL·E can compare the result of the diffusion decoding model and the image embedded by the CLIP engine. There are usually differences between the two pictures, so the content and style of both input photos are organically blended in the intermediate variants. This generates multiple images that convey the semantic information the user entered as the input caption (Ramesh, Dhariwal, et al. 2022b).

Humans find meaning in mages through the semantic details that provide subtlety, context, and emotional resonance (Leong and Mihalcea 2011). The images we read are replete with information, much of which describes these nuanced, indiscernible minutiae. OpenAI believes DALL·E's utilization of this two-stage sampling process confers a notable advantage by prioritizing the modeling of these high-level semantics that humans use for interpretation. While only a fraction of this information is instrumental in conferring visual coherence and object significance to an image, the CLIP embedding aptly captures a substantial proportion of this salient information (Ramesh, Dhariwal, et al. 2022b).

When a user enters a prompt into an AI image generator the task of translating their text parameters into an image is usually underspecified. A single caption corresponds to an infinitude of plausible pictures. While a user may have a specific idea in mind, the image generated is not uniquely determined by the language included in the request. Consider the prompt "gothic stained-glass window with medical line drawing of human torso." Depending on the orientation of the torso it may be necessary to include the window frame, although this detail is not mentioned explicitly in the prompt. Ramesh, Pavlov, et al., note that the DALL·E model has the ability to resolve underspecification in cases like setting, time of day, and artistic style (Ramesh, Pavlov, et al. 2021b). The result is it can draw the same object in a variety of situations. Through variable binding the model also has the capability to simultaneously control both the attributes and spatial relationships of multiple objects. If we were to alter our stained-glass prompt to, “gothic blue stained-glass window with medical line drawing of human torso wearing red and orange pants”, it must not only correctly compose the shape of the torso, but it must also accurately interpret and compose each pane of colored glass, while maintaining the association between “red and orange” and “pants”. All without intermixing the elements (ibid. 2021b).

This management of ambiguity and variable binding, combined with the efficient mapping of the concept space, affords DALL·E a diverse set of image generation capabilities. Developers have observed the model can create anthropomorphized versions of animals and objects by combining unrelated concepts in plausible ways. While one is unlikely to intentionally sit on a Persea americana, entering the prompt “an armchair in the shape of an avocado” transforms these two dissimilar items into renderings one might find in the furniture store (*see figure 39*).

Figure 39 Generation of anthropomorphized versions of animals and objects on DALL·E 2, images generated 22 March 2023

4.4 What is the current state of AI video generators?

The maturation of Artificial Intelligence (AI) has brought about significant advancements in the generation of images, text, and audio. However, the progression of Generative Video is advancing at a comparatively slower rate. This lag can be attributed to the complex challenges that arise from video generation. One of the primary obstacles is the sheer amount of processing required to produce a moving image. While applications like DALL·E may take 10-15 seconds to generate a single instance, creating smooth motion in video necessitates the creation of at least 24 images for every second of content. This number is not arbitrary; human perception of motion appears choppy when frame rates fall below 24 frames per second. Researchers have found that individuals process film running at 10-18 images per second as a sequence of individual pictures (Read and Meyer 2000, 24). During the 1920s, the motion picture industry wanted to minimize production costs due to the high cost of film stock.

Despite the fact that higher frame rates facilitated smoother motion, the industry settled on 24 frames per second as the optimal cadence. This rate was deemed the lowest viable frame rate that would still provide an immersive experience for moviegoers. Consequently, this industry standardized frame rate persists today.

The emulation of smooth motion in video requires an extensive number of images, leading to a significant increase in the computational requirements for generating even short video clips. In comparison to generating a still image in 10 seconds, generating a video clip may require constant processing on a neural network for up to 20 minutes. The increase in computation required for generating a five-second video clip is a factor of 120. Therefore, the production of Generative Video remains a challenging task due to the high computational cost required to create even a brief clip with smooth motion.

A second challenge is driven by the design of Generative AI's algorithm. These models are trained to produce patterns that solve narrow problems. Because there is no inherent structural incentive for an algorithm to generate alternative versions of an image once it arrives at a solution, resulting in mode collapse, the models have stochastic randomness baked into the code. The advantage of including random disruptions during generation is the model is always learning how to improve the output. The disadvantage is each generated instance is unique. Stochastic machine learning is a model that relies on randomness. Therefore, it is unable to guarantee accurate output. The impact of this stochastic randomness is the inability to repeat the creation of a specific image. AI creates what it wants.

For video generation, the inability to recreate images is a significant problem. The temporal nature of an object moving through a defined space requires the generated elements in the scene to appear identical from frame-to-frame. The appearance of movement in a video comes from each object's geometric change in relation to other objects. A video is a series of frames — still images — that are combined in a way that gives the illusion of movement. Elements coming closer appear larger over time, smaller if they are traveling away, and foreground objects move from side-to-side at a faster rate than those in the background. The coherence of movement is dependent on each object's appearance remaining largely unchanged from frame-to-frame, with temporal continuity driven by this geometric relationship changing over time. What breaks the illusion of coherent motion is when the shape and appearance of each object randomly changes from frame-to-frame. A viewer struggles to engage when backgrounds transform significantly, when objects appear in one frame and disappear in the next, when the orientation of the surfboard randomly changes direction, or when the surfing astronaut has three arms in frame 15 and then two in frame 16 (*see figure 40*). It becomes impossible for a viewer to submit to Coleridge's "willing suspension of disbelief" (Böcking 2008). They struggle to follow the action, and ultimately, the story. AI's difficulty with management of frame-to-frame temporal visual coherence results in visual chaos. In a New York Times interview, Runway AI CEO, Cristóbal Valenzuela, acknowledges that the weakness with these Generative Video models is temporal coherence, noting "The trick lies in training a model that understands the relationship and consistency between each frame" (Metz 2023, 4).

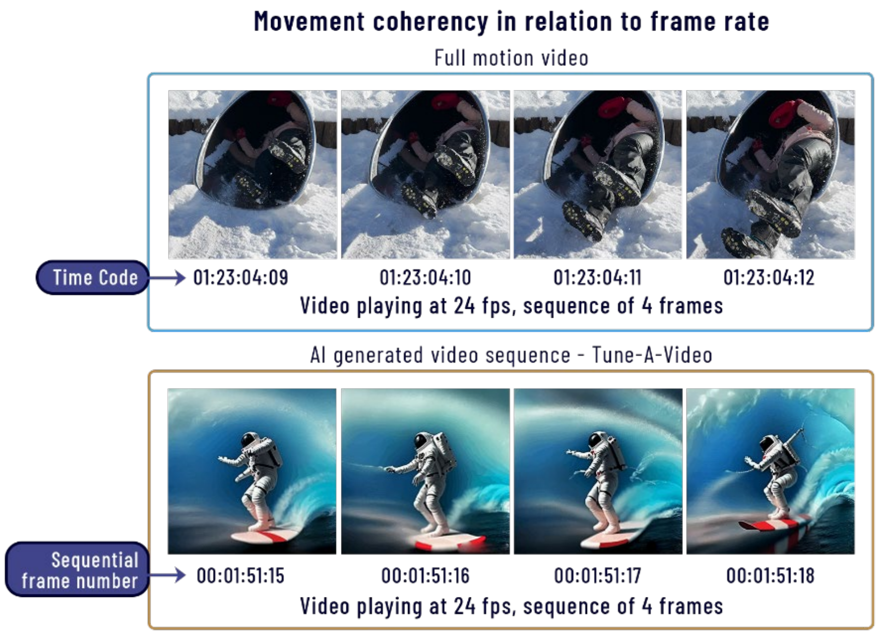


Figure 40 Comparison of sequential motion in video footage and the temporal incoherence of objects and backgrounds in AI generated video. Each example displays four frames in order as they appeared in footage.

While facilitating coherent synthetic animated video is progressing slowly, other applications of AI in video production are already appearing in media production via automation of mundane tasks. Some of these tasks fall within the scope of narrow AI. We have seen applications deployed for generating synthetic voice overs[5], synthetic audio localization through language translation[1], auto generation of captions via speech recognition[6], video upscaling[7], video masking and regeneration[8], and scene detection[2]. The most widely used applications are tasks not dependent on access to a neural network and can instead be executed locally on a user's machine. These tend to focus on computational

[5] Clipchamp video editor Text-to-speech generator. Here you can create a free voiceover from a range of voices in a variety of languages and accents for your videos. (ClipChamp 2023)
[6] Adobe PremierePro Speech to Text. Automatically generate transcripts and add captions to your videos to improve accessibility and boost engagement with Speech to Text (Adobe 2023).
[7] Topaz Video AI. Topaz Video AI focuses on deinterlacing, upscaling, and motion interpolation (Topaz Labs 2022).
[8] Content Aware Fill in Adobe After Effects. Remove unwanted objects such as automobiles, telephone poles, and people from your video scene. Leveraging Adobe's Sensei AI engine, the feature is temporally aware, so it algorithmically removes selected areas, analyzing frames over time to synthesize new pixels from previous of subsequent frames (Adobe 2022).

methods applied to manipulating the appearance and style of video. This is true particularly with applications like ClipChamp, aimed at novice video users, (Bar-Tal, et al. 2022). To date, the majority of the video image manipulation applications have focused on global style-transfer settings, where the appearance of an individual video clip is determined by a reference look or filter. This includes shifting a scene's time from day to night, compositing additional rain or snow into a scene, or replacing a face. Applications like text-to-motion graphics or easy-to-use text-to-3D modeling are on the near horizon. The more complex storytelling tasks of fully automated editing or unsupervised script production, two applications of Artificial Intelligence that will have a significant impact on video creation because of the emulation of a human point of view, appear to be much further from realization. Currently, much of the focus on AI video is on the development of text-to-video models. The progress in image generation models has created a likely foundation for synthetic video, but it lags behind the leaps we have witnessed in applications like Stable Diffusion and DALL·E. This is because of the complexity in modeling higher-dimensional video data, and the lack of large-scale datasets with high-quality text-video pairs (Singer, et al. 2022).

While the current failings of text-to-video generation prove the task is far more complex than generating static images, the technology is evolving rapidly. On a comparative arc of AI progress, the existing models are at the beginning of the curve, with major technology firms like Google, Meta, Adobe, and NVIDIA pouring hundreds of millions of dollars into development (Market.us 2023). They are willing to make this investment because within the next 7 years the anticipated valuation of AI Generated Video will exceed 2.2 billion dollars (ibid. 2023). If Video AI's evolutionary curve mirrors the rapid capability improvements in Generative Image AI, we are likely to see synthesized moving pictures with visual depth and fidelity, coherent object rendering, and seamless transitions between each scene's liminal edges within the next year. In simpler terms, synthetic video will match the quality and wonder of AI still images. To frame this development, it helps to explore the evolutionary timeline of AI video generation to offer a better understanding of how each approach is building on the evolution of CANs, CLIP and GPT.

One of the first instances of AI generated video was created by Alexander Reben. Titled *Deeply Artificial Trees* (2017), Reben's film featured a psychedelic composite of artist Bob Ross, overlayed on original source footage of an episode of *The Joy of Paining* (*figure 41*). These were early days when computing resources were limited, using comparatively primitive architecture. Reben created his film by feeding each video frame into a VGG or Google Deep Dream AI image model. To synthesize the voice, he fed an entire season of audio from *The Joy of Paining* into a WaveNet system (Tarantola 2017). The

WaveNet required a month of training to learn the speech patterns of Ross. One indication of the rapid advancement in neural modeling is by contrasting the 30 days required for Reben to train his audio on the WaveNet, with Microsoft's VALL-E system six years later – a model that can accomplish the same task in as little as three seconds (Wang, et al. 2023). The rather hallucinatory nature of Reben's film was a result of Google's Deep Dream algorithm. To generate an image the model would read an individual frame of video, attempting to recognize objects in the frame. It would then generate what it believed it had seen and overlay it onto the frame, in essence replacing the original object with the AI interpretation of what it saw. The temporal continuity from frame-to-frame was somewhat successful as the model was reasonably consistent in reading and placing specific objects. The film still suffers from significant strobing and pulsing in shapes and patterns throughout the film, and Ross' head often transforms into random objects, shifting from a lion to a bird to a cabbage, within a span of only a second.

Figure 41 Deeply Artificial Trees, 2017 Image Courtesy Alex Reben.

Five years later Meta unveiled an AI video model, Make-A-Video, that could generate scenes with instructions from a text prompt (Meta 2022). In the paper, "Make-A-Video: Text-to-Video Generation without Text-Video Data", Uriel Singer and a team of Meta researchers discuss their approach to animating a single photograph into a brief 3-second clip (Singer, et al. 2022). Their model was predicated on leveraging existing advancements in text-to-picture image generation, with the authors specifically referencing DALL·E in their paper. Because training an algorithm on a video dataset is inherently complex, the developers chose to apply existing image deep learning models, GAN, and diffusion image generation architecture to generate video from prompt images.

Like all generative models, video algorithms require a very large dataset for deep learning. In this case, text/video pairs rather than text/image pairs. This requirement is problematic because unlike the process of training an image algorithm one photograph or one painting at a time, training a video model implicitly requires incremental learning. It is not mapping one image to an instance of text, rather is it mapping an entire scene one frame at a time -- for each instance of text. The consequence is the compute required for this training grows exponentially. One can train a CLIP image model on a relatively small set of 12 million pictures. If you attempt to train a video model with the same number of assets, you will be processing 12 million video files with multiple frames in each video clip. Your compute could easily expand to 1.2 billion frames[9].

The second complication is, unlike images scraped from the internet, video files are rarely hosted on individual websites. The consequence is video files used for training are not as freely available (Singer, et al. 2022). Most online videos we see are embedded as a video player, originating from 3rd party services like YouTube, Brightcove, Vimeo, and Adobe. For data aggregators like Common Crawl, accessing these repositories becomes far more difficult. This is because most videos are either shielded behind security features or do not have the required descriptive alt-text association for each file. Of the data sets that are available, most are controlled by hosting services. Google's dataset, YouTube-8M, dates from 2019 and holds 237,000 segments of video (Google 2019). The segments are human-verified and categorized into 1,000 classes. They were culled from 6.1 million videos, totaling 350,000 hours of content. While 237,000 organized segments are helpful for broad training, the size pales in comparison to the 650 million image-text pairs used to train DALL·E. Additionally, it has been nearly four years since Google Research has updated the collection.

Because of the potential training compute costs and limited access to video datasets, Meta chose instead to use image synthesis data (still images trained with captions) and unlabeled video training data for deep learning (Singer, et al. 2022). The videos generated by Make-A-Video are brief, averaging between 30 and 75 frames. To synthesize a clip the model creates the first frame from a text prompt. It then animates movement by applying its understanding of potential object location over time and space. As it processes each frame it guesses what should happen next, attempting to maintain image and temporal coherence, generating a brief scene in motion.

In the same year, GPU developer NVDIA released a model that generated higher resolution videos of longer duration. Their LongVideoGAN algorithm has the capacity to create 256x256 pixel sized clips of continuous motion, which is then reprocessed in a second pass to create high fidelity coherent scenes (Brooks, et al. 2022b). The videos generated suffer somewhat from the temporal artifacts described earlier, with objects wiggling on the screen, and side-to-side pulsations of misshapen features of people and animals. The significance of the NVDIA model was improved object motion and the ability to smoothly change the camera's viewpoint during the scene. Prior to this release, many synthetic AI videos suffered from an inability to maintain consistent and plausible motion over time. They were

[9] Assuming average clip length is ≅ 100 frames, or 4 seconds at 24fps. 12,000,000 x 100 = 1,200,000,000

either locked into a single camera perspective or portrayed a single action, relying on inductive biases to meet the need of maintaining temporal consistency. The composition within a video would be unchanged, dictated by the initial compositional framework. This limitation also impacted the physics of objects, where movement and actions appeared surreal. Feet would float rather than step, waves lapped awkwardly, or a ball's bounce did not match its inertia. Despite the temporal incoherence producing occasional frame-to-frame visual anomalies, the LongVideoGAN model displayed improved object motion and scene continuity.

Of particular note was the model's ability to transition from one scene's action into the next. Objects in previous examples of longer duration generated videos would often morph unrealistically between scenes. To address the need for consistency the NVDIA team trained their model using longer videos, prioritizing an understanding of motion over time (ibid. 2022b). Because of the algorithm's demanding deep learning phase, they initially trained it with long low-resolution video files, followed by shorter videos at high resolution. This emphasis on long-term temporal dynamics produced media files with fewer seams from scene-to-scene.

Another NVDIA model released in 2022 addressed AI video's potential to augment existing video through the manipulation of objects in the frame. Text2Live is a method to alter objects in a specified video using a text prompt (Bar-Tal, et al. 2022). Omer Bar-Tal and a team of developers from NVDIA and the Weizmann Institute of Science created a zero-shot model where a user can describe intended manipulations of elements in a scene. This manipulation can be an object's texture (like the addition smoke and fire or changing the season) all while maintaining a scene's semantically meaningful context. Unlike the process of generating all the visual elements in a scene, the method utilized in the Make-A-Video or LongVideoGAN models, this approach is closer to Reben's *Deeply Artificial Trees* where generated elements are superimposed over the original footage. To add smoke to the scene in figure 40 of a man smoking a cigar, a user first specifies the intended appearance of the target and the object/region to be edited (*image 42*). The model maps the size and location of the man's face, generates a puff of smoke floating out of his mouth, and then composites the two images together. They write in their paper, "Text2LIVE: Text-Driven Layered Image and Video Editing" (2022), this method automatically locates the region and synthesizes realistic, high-quality texture that combines naturally with the original image, performing the augmentation in what they describe as "a semantically-aware manner" (ibid. 2022, p 2).

Like DALL·E, Text2Live was trained using a CLIP model, leveraging 400 million text-image examples from the ViT-B/32 pretrained model (ibid. 2022). They chose to use CLIP because of their

Figure 42 Text2Live Compositing model. Adapted from: Text2LIVE: Text-Driven Layered Image and Video Editing (2022)

belief this path yields significant visual and textual benefits. The significance of the advances with the Text2Live model are twofold. The first is it capability to extract elements and styles from objects in the training dataset and apply this data as texture maps. Textures like fur, stained glass, or smoke, are generated and applied as geometric shapes or objects in the original scene. The second advancement is their choice of file color space, 32-bit color[10]. Unlike most synthetic video algorithms that generate moving images by outputting every pixel in a scene, Text2Live creates a separate layer comprising only the elements that will be added to the frame. This approach allows them to map the generated elements to specific locations within the frame. In previous models, adding a separate a layer of generated media to an existing video file would require a user to mask out (to hide) the areas that would not be included in the finale file. Using 32-bit color, or RGBA, the developers have created a production path that eliminates the additional work required of applying a user-provided edit mask.

The process of compositing a separate layer of generated media with a real-world video source file is inherently difficult. To embellish the source, a user inputs the video file into the Text2Live model, initiating a process of decomposing the video stream by pre-training a set of 2D Neural Layered Atlases

[10] The dimension of a high-definition video frame is measured in pixels, with each frame containing a total of 2,073,600 pixels, at 1,920 pixels wide by 1,080 pixels tall. The hue, saturation, and luminance for each pixel are determined by mixing three color channels: red, blue, and green (RGB), with a numeric value assigned for each channel. The RGB number for each pixel is stored in the file using 8 bits of memory for each of the three channels. Most video images contain only 24 bits of color information per pixel (3 channels x 8 bits = 24 bits). When compositing additional elements on a video, the composited layer file must also include opacity instructions. This is an 8-bit number with a value between 0 and 100, which indicates whether a pixel is transparent or opaque. The opacity information, called an alpha channel, occupies an additional 8 bits of information, for a total of 32 bits of information per pixel. As this additional 8 bits of information increases the total file size, most image and video file formats do not include alpha information. This is because alpha data is not leveraged with streaming and broadcast media, and larger files are more difficult to process. In the cigar smoke example, most of the pixels will have an alpha channel value of 0, indicating those pixels are transparent and will not appear in the image. The pixels used for the smoke will be opaque, with values in the alpha channel approaching 100. RGBA formats are most often used in image compositing, text overlays, special effects, and motion graphics.

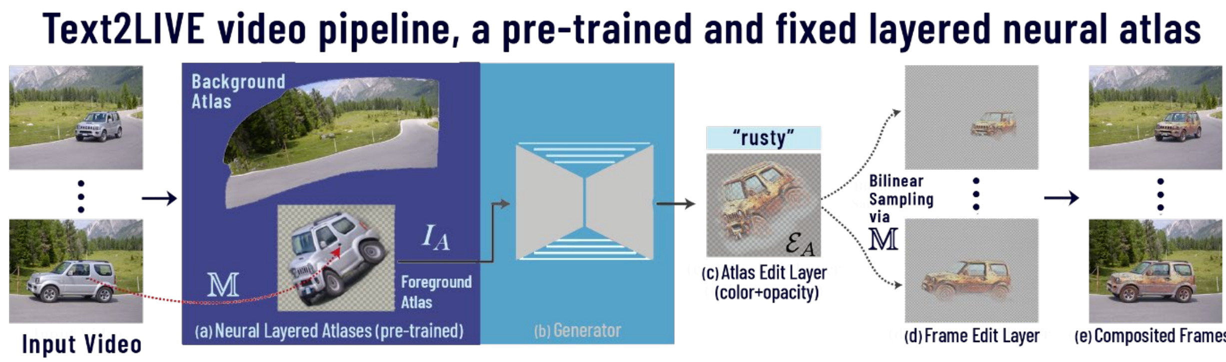


Figure 43 Text2Live video pipeline with Pre-trained Neural Layered Atlas. Adapted from:
Text2LIVE: Text-Driven Layered Image and Video Editing *(2022)*

(*see figure 43*). Each atlas will be treated as a unified continuous 2D image, representing either a foreground object or the video's background. By mapping everything prior to image generation the AI identifies specific objects, determines the geometric plane for each object, understands the relationships of each object in a synthesized 3D space (what will pass in front or behind), and identifies the edges. This information will be used as templates for implementing the final composite.

Next, the model generates synthesized video by combining data from the user's prompt and a feedback loop from the atlases. The prompts describe not only the final edited image, but can also target an applied effect (e.g., rust, stained glass, fire). This approach affords a user the ability to not only specify the overarching style and appearance of a scene, but also identify effects for the specified targeted objects. The file generated from this process is called the Atlas Edit Layer (the RGBA file), and it will be combined with the original input video to produce a new augmented file. Bar-Tal, et al. believe there is inherent flexibility with this approach, adding efficiency and nuance through the application of training information that is readily available within the source video file (Bar-Tal, et al. 2022).

Stable Diffusion co-creator, Runway, has also released a model that that explicitly applies source file information for training. Called Gen-1, their model uses the original media's basic composition, movement, and geometry, transforming the video by applying the requested look or style. In the paper, "Structure and Content-Guided Video Synthesis with Diffusion Models", authors Esser, Chui, et al. write that Gen-1 is a content-guided video diffusion model that edits videos based on visual or textual descriptions of the desired output (Esser, Chiu, et al. 2023). Examples of this style application would be transforming a shot of people on the street into Claymation puppets, altering martial arts practice into a line drawing, or a turning a commuter into an anime character (*see figure 44*).

Where previous video generation models suffered from object and background temporal inconsistency, resulting in significant frame-to-frame errors, Gen-1 can generate realistic and consistent video by applying motion data from the source. Most generative models are recursive, which results in scene generation where the movement feels chaotic. This is because the models know where they have been but not where they are going. The movement in the frame is predicated on guessing an object's next position in the subsequent frame based on where it was in previous frames. It has a rudder but no North star. By applying the motion from the source footage as a motion framework, the model can generate natural movement of objects in the frame. The data culled from the source footage guides the model as it decides an object's size, position, geometry, and relationship with other objects. Where models like Meta's Make-a-video and Google's Phenaki (Villegas, et al. 2022) are limited to brief 2-4 second clips, and the Text2Live application is limited to generating only specific elements within a frame, Runway's Gen-1 has the potential to produce much longer videos. This is because of the training and temporal efficiency derived from existing footage's data. In a blog post the Runway research team writes that they believe their model can consistently synthesize new videos by combining the composition and style from a text prompt with the structure of the source video clip. They add, "It's like filming something new, without filming anything at all" (Runway Research 2023).

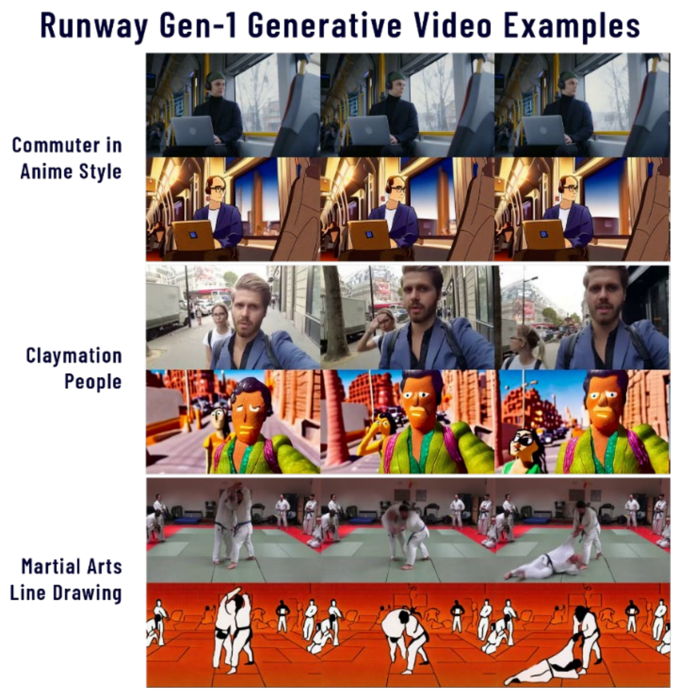


Figure 44 Video examples of source video footage used to train the application of a generated style. Adapted from original source: Structure and Content-Guided Video Synthesis with Diffusion Models, (Esser, Chiu, et al. 2023)

Runway has been developing AI-powered video-editing software since 2018, including mainstream social media video filters for TikTok and powerful effects generation tools for television and film studios (Heaven 2023). Runway's technology is often developed in close collaboration with a community of video producers and filmmakers, including the team behind the visual-effects-heavy film, *Everything Everywhere All at Once* (Tangcay 2023). It is a natural fit as the film industry has been editing and manipulating film for more than a century. Runway plans to incorporate this generative model with

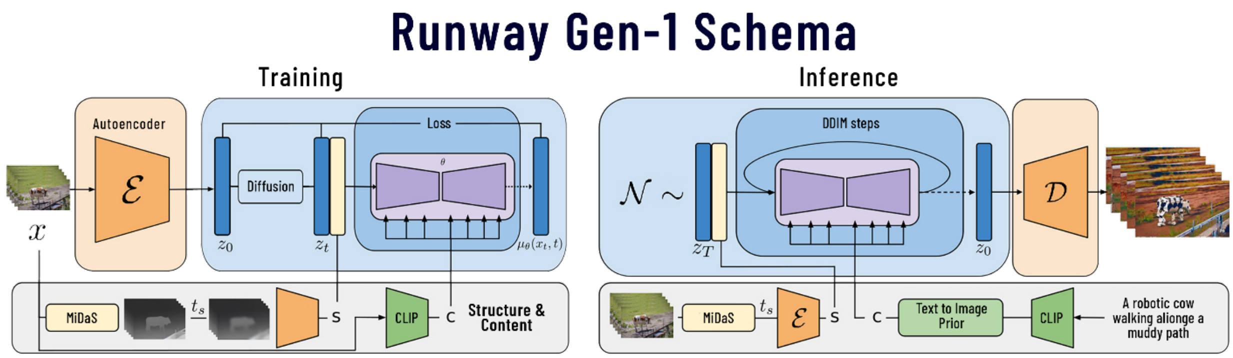


Figure 45 Gen-1 video generation schema. Adapted from original source:
Structure and Content-Guided Video Synthesis with Diffusion Models, (Esser, Chiu, et al. 2023)

other video effects tools, believing it will speed up the workflow of professional artists (Metz 2023). This real-world testing has put their model on a fast-track for widespread adoption, with Runway releasing an improved version for testing, called Gen-2, only 4 months after the launch of Gen-1. They are betting that a user will want to generate videos, complete with music and dialogue, simply by typing a prompt on a computer screen (ibid. 2023).

The two-step process used in the Gen-1 model is an initial training stage followed by the application of that knowledge in an inference stage. Similar to Text2Live, the model is learning a specific clip's content and structure (*see figure 45*). The content phase defines features that describe the video's appearance and semantics (Esser, et al. 2023). Here the model is learning aspects like the colors, styles and lighting of objects. The data gleaned from the structure phase includes information about every element's geometric and dynamic movement. The model maps the shape and location of each subject, as well as their temporal changes through the duration of the clip (ibid. 2023).

During training, input videos are encoded with a fixed encoder and then diffused. Each video's structure representation is extracted by encoding depth maps via the MiDaS engine, and a content representation is encoded with a CLIP model by selecting only one of the video's frames. The model then learns to reverse the diffusion process in the latent space, with input provided from the MiDaS and CLIP subroutines. To generate the new video clip, the user enters a text prompt describing their video file's intended stylistic modifications. The text prompt is converted from a CLIP text embedding to image embeddings via a prior. The input video (the scene to be modified) follows the same path as the initial training videos of encoding, diffusion, depth maps, and CLIP content representation. Through the application of the instructions from the prior and the structural and content information from the MiDaS, the diffusion model is capable of generating temporally coherent videos. The authors state that their goal is to edit a video while retaining its structure (ibid. 2023, 3).

Runway's approach extends existing image diffusion models to video generation through the application of temporal data in pre-trained image models, with the addition of training jointly on images and video. The research team at Runway believe that by training their algorithm on increasingly large amounts of data, they can rapidly improve the fidelity of clip generation (Metz 2023). It is difficult to describe the current output from the model as the files are neither a photo nor cartoon. It is an amalgamation of pixels that blend together to produce a slightly surreal yet uncannily realistic video. The current output appears to produce the best results when a prompt specifies something with small movements or action, like a steaming cup of coffee or rainy day in a forest. Because the model is both structure-aware and content-aware, the editing is performed entirely at the time of inference. This eliminates additional pre-processing, compositing via an image overlay layer, and per-video training. The approach affords a user the ability to control how strongly the model adheres to temporal consistency in generated clips.

Gen-2, Text2Live, LongVideoGAN, and Make-A-Video are four examples of the rapid advancement of AI video generation. Other recent models include Google's video diffusion model Dreamix (Molad, et al. 2023), and Adobe's GANgealing (Peebles, et al. 2022). Taken together they provide evidence of a clear progression in capability, with each representing a different approach to solving similar challenges. Still, none of them demonstrate generative fidelity comparable to what we have seen in large language and still image generative models. The advancements in video that we have witnessed over the last two years imply that synthetic video's place on the AI evolutionary curve is still in the early days. The pace of development points to technical advancements in visual depth and fidelity, coherent object rendering, and seamless transitions between each scene's liminal edges within the next year.

4.5 Where do we go from here?

Alan Thompson predicts that, "Text-to-video is going to be huge in 2023" (A. D. Thompson 2022b, 10:50). He points to the potential combination of generative video models with automatic speech recognition systems like OpenAI's Whisper (Radford, Kim, et al. 2022). He envisions a marriage of advanced text-to-video models and language recognition, potentially automating the entire process of video creation. "Text-to-video models will go some way to setting up your next imagined film or TV series in an instant, and then your entire virtual reality environment" (A. D. Thompson 2022b).

While Thompson's hyperbolic AI prognostications may seem more rooted in an imagined potential than near-term reality, it is likely that with each advancement in AI Generative Video we are a step closer towards an unspoken goal of automated storytelling. What is unclear is who is setting this goal? Is the desired objective total automation for media production? His enthusiasm leads one to the question, what is the destination for AI motion media creation?

Video generation is but one component of video storytelling. It is a single voice in a chorus of narratives woven together to craft a tale. Generating a successful video using only a text prompt will require more than large language models, diffusion, CLIP, and zero-shot music generation. It will require the nuanced integration of *all* the voices of storytelling – image, music, plot, and environmental sound. The generated narrative will need conflict, heroes, resolution and a "moral to the story". It will need the arc of a beginning, middle and end. The thousands of choices made by a storyteller reflects not just their point of view, but their humanity. These are choices measured in increments of frames, decibels, and words. What seems to be lost in this enthusiastic march of AI progress is how generative models will fit into our personal demonstrations of creativity. While we are seeing technical advancements in generated video, what will the resulting video communicate to a viewer? Will the story's message reflect the creator's point of view? The developer's? The algorithm's? Answers to these types of rhetorical questions will remain elusive in the near term. Still, there is merit to posing them as they will frame how we choose to adopt the technology. The answers will become a metric for measuring what we gain and what we may lose.

A likely consequence of incorporating AI into our lives is we will be outsourcing much of the mundane minutiae of our day. It is unlikely we will miss the tasks of answering email, creating the perfect image for a sales brochure, or assembling a shopping list. Putting those jobs in the rearview mirror will be easy for most. The struggle may come from drawing the line between delegation and submission. How do we determine if eliminating these moments of mundane minutiae ultimately shifts our focus from the role of creator to supervisor? In act of expressing ourselves through words, will AI lead us to transition from writers to editors?

The distinction between writing and editing can be likened to the distinction between writing and reading. Reading closes gaps in our knowledge by expanding our understanding, but it does so using language crafted by others. The words, sentences, narratives, and arguments belong to someone else. Editing also starts with language produced by another individual or machine. Writing differs from reading and editing because it offers a unique avenue for understanding. To some extent, writing is a form of editing because the structure of language and its conventions are passed down through culture. Yet to rely on AI as the primary avenue for generating representations of our thoughts, feelings and discoveries has the potential to limit what we know of ourselves. To delegate the task and forego the act of writing is to exist solely within the confines of language created by others. It is to rely on edited, revised, and embellished language rather than constructing individual meaning through the written word. Language serves as a medium for human beings to comprehend ourselves and the world.

As a storyteller, I believe the stories I create are vehicles for sharing how I comprehend the world. My medium is video because the process of combining the personal voices of image, words, sound, and music, weaving them into a tale, is profoundly satisfying. The combinatory act of creation reveals connections that were previously unseen for me. I want to be responsible for every element and every decision in the videos I produce. These decisions are shaped by a lifetime of experience and knowledge, inspired by the work of others but not copied directly. The edits I make are formed by a cascade of tiny catastrophes, experiments that fail, expertise gained though trial-and-error, with the lessons learned applied going forward. These are the experiences of someone who has invested decades of struggle towards the acquisition of capability in a craft. Perhaps it is vanity, but I find satisfaction in the struggle and the accomplishment that comes with ability.

However, if one never learns how to edit a video, the ins and outs, the little quirks of editing, if one delegates the act of writing to a chatbot, if one never discovers how to compose a photograph, will they possess the abilities that come from experience? The likely outcome is the value of these skills will be depreciated in the same way lacemakers saw their skills devalued by automation. When a human is inspired by the artwork of others it becomes applied knowledge, wisdom incorporated into their creative point of view. When a Creative Adversarial Network performs the same task, but at an unfathomable scale, the probable result will be derivative work that brings the cost of creation down to near zero. The evidence left in the wake of many AI developer's efforts, those who are leading the charge in generative AI, often demonstrates a lack of understanding about consequential disruptions to existing ecosystems. They have proven that they are adept at applying algorithms to solve difficult empirical problems. What is less clear is whether they understand that not everything requires an empirical solution. Bret Devereaux believes just because they understand the complex task of computer programming, they assume "that all other complicated things must be lesser in complexity and naturally lower in the hierarchy of reality. Nails easily driven by the hammer that they have created" (Devereaux 2023).

We are facing disruptive consequences that are likely to impact creativity, personal expression, and the ways we reflect our values to ourselves. There is the potential for AI to give voice to those who have neither the knowledge nor capability to master complex creative tasks. There is the possibility we will see AI creator communities blossom, just like we have witnessed YouTube's empowerment of video storytellers, fostering communities among underrepresented groups. There is the potential for AI to extend human creativity by affording exploration through the recombination of existing ideas. There is the potential to streamline the processes associated with creativity, freeing us from the mundane so we can focus on the complex. There is the possibility for artists to create at scale through automated brainstorming, keeping instances of merit and discarding the rest.

Conversely, the development of Artificial Intelligence is likely to result in a devaluation of effort. Like lamplighters, wheelwrights and shoemakers who saw technological innovation destroy their livelihoods, accountants, administrative assistants, and video editors may face a job market where their supply far exceeds demand. A 2023 Goldman Sachs report estimates that AI will impact 300 million full-time jobs globally, with 25% of the current workforce substituted by AI, and two-thirds seeing some degree of automation to their jobs (Briggs and Kodnani 2023). Clearly the impact of AI will be huge. The challenge associated with unraveling the potential outcomes from AI implementation are not only the scope and scale of deployment across global industries, but the unforeseen and unintended consequences of change.

Chapter 5

The Storytelling Model

At the outset of this paper, we established that both human and algorithmic creativity are defined as the ability to generate new, valuable, and surprising ideas in an unfamiliar or unexpected manner. We also discussed how we appreciate creative demonstrations when the resulting media reflects both the aesthetic and emotional facets of human intelligence. In Chapter Two, we examined how the influx of automated creative paths is threatening to displace artists and the importance of developing tools that allow artists to maintain their creative agency, distinguishing their work from products generated by machines. Chapter Three surveyed the last 80 years of AI's stratospheric development and the contemporary acceleration of its capabilities. Chapter Four outlined how generative models can synthesize text, music, voice, and images, and how we are on the verge of generating full-motion video. These observations about creativity, creator agency, and AI development lay the foundation for the argument that video storytelling is an area ripe for AI invention. However, creating a video story that engages a viewer is much more than simply generating its individual components. It requires integrating all the elements of storytelling—images, words, characters, plot, and sound—into a coherent whole. How realistic is it to develop an algorithm capable of evaluating source footage and generating a suggested story?

5.1 The ubiquity of digital video

We are surrounded by digital video. Creators upload 720,000 hours of video to YouTube every day (Mohsin 2021) (Hale 2019) (Ceci 2023). Amazon's streaming channel, Twitch, averaged 2.8 million concurrent viewers in 2022 (TwitchTracker 2023). Estimates are that in 2023, 82% of global internet traffic will come from streaming videos and downloads (Cisco 2020). Video has become the world's lingua franca, a common language of the human experience. The ubiquity of these always-available and ever-present digital video streams sets the expectation that the information one seeks is available digitally, online, and streaming.

As consumers we know what we like. We quickly measure a video's merit against our personal expectations of information, entertainment, and engagement. Being steeped in video, we have come to see the medium as a means of personal expression, made possible by a camera we always carry with us — the cell phone.

Still, most people find that video creation is difficult (Rodgers and Dhonnchadha 2018). Creators struggle because most have been trained as consumers, not as video storytellers. Crafting a coherent narrative can feel opaque as video production requires a functional literacy in a variety of disciplines. Beginning creators face three primary challenges – expressing their voice as a storyteller, using the camera effectively, and navigating editing's steep learning curve. A creator needs to understand the architecture of storytelling, identifying their story's expected outcome and defining what will be required for the story's beginning, middle and end.

They need to leverage creative tools to generate the video assets required for crafting a visual narrative. As an instructor of documentary production, I have seen my students struggle with the camera. They often feel overwhelmed as they disentangle the syntax of sequencing in an effort to articulate a story's required compositional visual grammar. When shooting is completed, they will face the learning curve of a computer-based editing application, working to assemble their parts into a coherent whole. All too often I have witnessed disappointment when a creator's final product did not meet their expectations, nor the expectations of their intended viewers.

What many creators don't grasp is that most genres of online video follow specific patterns, and when measured by audience engagement and number of viewers, the most successful online videos are crafted within the narrative scaffolding of a specific genre (Chi 2020). For example, many successful how-to videos use a formula of:

a) setting up a problem or goal with a teaser of the outcome,
b) identifying the task's required resources,

c) step-by-step instructions,
d) revealing the completed project at the end.

A video that recaps an event (political, sporting, community, or news) will:

a) give context about significance,
b) set the stage and location,
c) select important moments (usually presenting them in chronological order),
d) show the climax of the event,
e) end with reaction or analysis.

Even online "vloggers" leverage what has been described as playwriting's traditional 3-act or 5-act structure (Berman and Slobin 1994).

As we have explored through several examples, pattern recognition is a strength of AI. An Artificial Intelligence model is well suited to learn the patterns of video storytelling. Videos are comprised of well-defined sequences that are layered and repeated, nested inside each other like a Matryoshka doll. These are patterns that can be learned one layer at a time, stored as a sequence of events, and applied in reverse order.

Most digital videos fall within a specific genre. Each genre has a basic structure that has evolved through tens of thousands of iterations, with both creators and users measuring what is most effective. Like a play or film, a video is structured into acts. A viewer will rarely notice this construct because when a story transitions from one act to the next a filmmaker will strive to make these transitions seamless. They are endeavoring to keep the audience engaged. Within each act are scenes, smaller stories that stand on their own with a clear beginning, middle and end. Each scene comprises a structured sequence of images and information, crafted by the storyteller and captured using camera angles and shot composition that communicate specific meanings. Decisions like where the storyteller places the subject in the frame, the size of the subject, the elevation of the camera, and the lighting, all provide information that viewers understand subconsciously (Mercado 2022). The order of shots presented within a scene has a very specific grammar. This order provides context, a sense of place and time, provides insights about the subject's feelings, and ultimately delivers the scene's resolution. The sequence of images that make up a scene, the sequence of scenes that make up an act, the sequence of acts that make up a story, all fit general patterns for each genre. An AI model has the potential to replicate the actions and decisions followed by human filmmakers when they assemble a story. Iteratively moving from crafting the micro patterns in a scene to the macro patterns of a genre in a sequence of story assembly.

Recent work utilizing AI for video creation show promise in achieving this goal in the near-term. Ali Javed has led two interesting studies exploring the feasibility of developing an AI engine that automatically identifies key events in source footage. The first is a model that analyzes both video and audio to identify key moments (Javed, Irtaza, et al. 2019). The second builds on that work to then segment the selected moments into unique clips (A. Javed, K. Malik, et al. 2020). Hansa Shingrakhia and Hetal Patel have also developed a hybrid machine learning approach for creating highlight summarizations of cricket sporting matches (Shingrakhia and Patel 2021). Developers world-wide are beginning to tackle the problems associated with using machine learning to understand video content (Hernandez, Bulitko and Hilaire 2014), ultimately with the ability to classify it, segment it, and apply these patterns for creating AI algorithms that generate stories.

I propose a storytelling model that will leverage machine learning to recognize patterns in a specific genre of video stories (patterns of story structure, narrative content, and image sequencing). It will apply this knowledge to raw footage input by a user, with the goal of outputting instructions for creating a video that matches the structure of the chosen story type. I will explore the individual elements of my proposed model in detail, but on a high level, the model is designed in a two-phase approach, Training phase and Generation phase.

As an overview, in the deep learning phase, the model will learn to recognize story types and specific categories of images and sounds by analyzing contemporary examples of stories within a specific genre using a dataset of alt-text/video examples, building a CLIP library of image types, story, act, and scene construction (*see figure 46*). In the story generation phase, a user will input source video for analysis and enter a prompt describing the desired outcome (*see figure 47*). In broad terms, after indexing the available media, the model will generate a script via a large language model, leveraging keywords and source audio gleaned from the input source. The generated script will be sent to a synthetic voice generator to create the story's narration files. The model will use the index of source footage to compile a recommended sequence of shots, identifying the beginning and end for each clip. It will then find keyword matches by

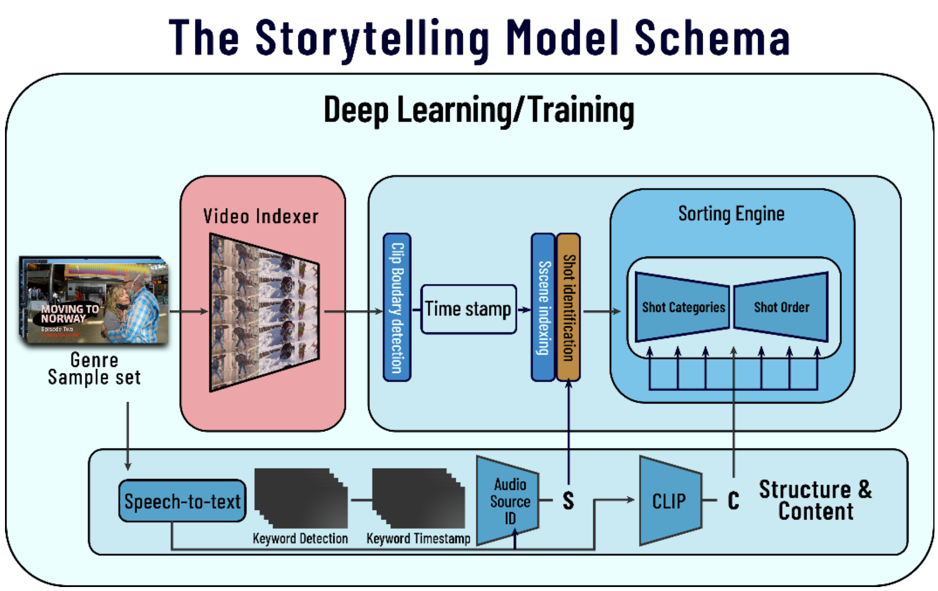


Figure 46 Training schema of proposed video production storytelling model.

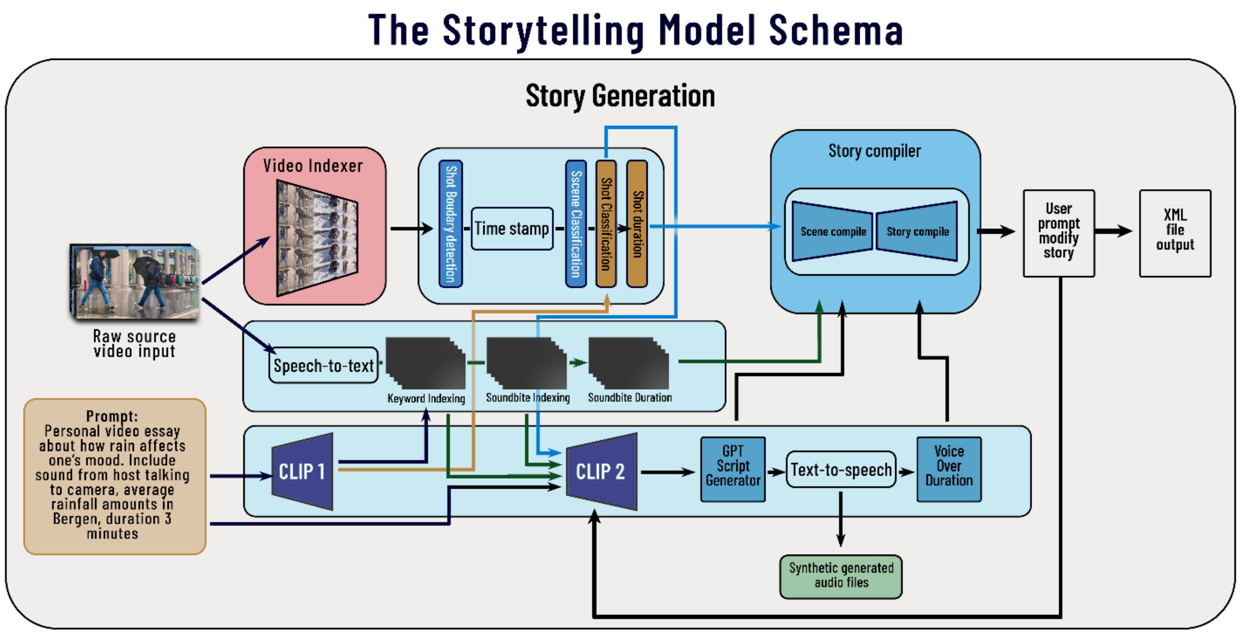


Figure 47 Story generation schema of proposed video production model.

comparing the available indexed footage with the script, aligning the footage with the narration files, and adjusting the duration of the source footage to match the length of the narration. The model will execute this process in stages, first compiling individual scenes, then compiling scenes into acts, and finally acts into the final video. This sequence of generated events, the beginning and end of each shot and the beginning and end of each audio file, will be compiled into a metadata string. Before outputting the final list, a user's story will be presented as a list in a user interface. This will afford a user the opportunity to adjust the generated story, and/or refine the initial prompt for regeneration, prior to final output. The model will output the generated story as an edit decision list, a string of instructions for assembling the story in third-party video editing software. This list will be structured using a common text-based format called Extensible Markup Language (XML). XML is an ideal format because it is a universal file accepted by most video editing platforms.

This proposal is predicated on certain assumptions, based on both ongoing Artificial Intelligence research and common practices in media production:

- Artificial Intelligence and Machine Learning have advanced to a point where an algorithm can analyze existing video examples to recognize patterns in story structures.
- These algorithms can also identify important moments in source footage by applying what it has learned from story examples.
- This engine would use raw source footage input by a user to generate a suggested video story. The algorithm would apply what had been learned

> from previous training about genres, images, sounds and story structure. The model will not compile the source footage into a completed video, but instead output an XML edit decision list.

There are many inherent challenges with this proposal. For example, the mechanics of compiling story assets in a tiered structure of scene, act, and story has the potential to pose problems during the allocation of a user's available assets found in their source footage. It will likely prove difficult to train the model as it works to find solutions using limited resources. An expected problematic instance would be when the model identifies a specific shot within a scene as important to the story, yet it is unavailable in the footage. While these sorts of technical hurdles could ultimately doom the model to failure, I believe the greatest challenge will be determining if a working model acquires the capability to generate coherent stories that engage. In other words, will the model learn to be a storyteller?

The potential of AI to create compelling video stories is likely to be met with both skepticism and enthusiasm. Some will argue that the technical challenges outweigh the value of the result. Others are likely to envision possibilities of creation, opening new avenues for creators to discover their voice. To better understand this complexity, it is important to examine the power of storytelling itself.

Jonathan Gottschall observed that everyone does story in one form or another (Gottschall 2012). Stories have tremendous power over us. Some theories postulate that they act as "social glue", provide escapist pleasure, afford practice for social life, and provide cognitive play (ibid. 2012) (Boyd 2009). They can transport us, resonating within as they carry us on an emotional journey. They can mobilize action or spark new interests (Chu and Roy 2017), or they can fall flat, boring us through predictability or, even worse, confusing us with a chaotic narrative.

The crux of our capacity to use stories as avenues of communication lies in the symbiosis of shared emotional experiences. Narratives unravel through distinct emotional arcs that coalesce into resonant patterns of significance. Douglas Massey writes that emotional narratives are a convincing medium for explaining the world we inhabit, enforcing societal norms, and giving meaning to our existence (Massey 2002). In a 2016 study, researchers leveraged a massive dataset to study the evolution of culture through stories. Using Project Gutenberg's fiction collection, they classified six core emotional arcs which form the essential building blocks of complex emotional trajectories. They classified these common story tropes as 'Rags to riches' (rise), 'Tragedy', or 'Riches to rags' (fall), 'Man in a hole' (fall-rise), 'Icarus' (rise-fall), 'Cinderella' (rise-fall-rise), and 'Oedipus' (fall-rise-fall) (Reagan, et al. 2016). This classification of classic story arcs points to a viewer's expectation that as it begins, they know how a story will unfold. At least generally. Stories are emotional arcs that meet and exceed our expectations. The joy of immersing ourselves into a story is often a balance of anticipation and surprise.

We have a basic idea how the plot will unfold and hope the way it unfolds includes elements of surprise. For a viewer, reader, or listener, there is implicit trust that the creator is in control.

Stories are innately human. Before investing the time required to watch a video, a viewer carries an expectation of knowledge to be gained, or laughter, or vicarious travel, or outrage, or a myriad of other emotions. These stories unfurl in an ever-building sequence of choices by the storyteller, resolving with a "moral to the story", punchline, or completed task. The sum total of effective video stories is more than the parts of information contained within. These parts add up to reflections of our shared emotional experiences. Which brings us back to the question, can an AI model learn to be a storyteller when it is essentially a math equation?

At the beginning of this thesis, I asked: if the art of storytelling currently reflects our collective human experience, what shall become of our sense of self in the face of AI-generated content? If narratives serve to orient us within the fabric of society, what are the implications when these stories no longer embody the tenets of human exploration, but instead are an aggregation of the thousands of stories that have come before? Will the artifacts we leave behind reflect humanity, or a homogenized synthesis of humanity -- an algorithmic assemblage that represents the mean of a universally popular dataset? These sorts of rhetorical questions are little more than suppositions about an obscure future. Yet there is relevance in keeping these questions in the forefront of our mind as we confront the impact of synthetic storytelling.

5.2 Testing the current tools

For the last year I have used a wide variety of platforms to gain a better understanding of the AI media production tools currently available. Because this is a dynamic and rapidly changing landscape, many of the experiences I share here are likely to be out of date the moment I describe them -- made obsolete by revisions, new releases, and expanded capabilities. Still, there is merit in discussing current applications as a window into existing technology that can be incorporated into the Storytelling Model. I will touch on a few that represent practical applications of AI that focus on solving common video production problems. Specifically, I will discuss Canva, Adobe and Runway.

Beginning in 2020, the first wave of AI in media production landed in two broad categories – lightweight, thin, gimmicky apps with little competitive differentiation, or niche applications aimed at video post-production. Many of the lightweight applications were ideas rushed to market, similar to the way hundreds of apps were rolled out after the release of the first iPhone. However, researchers Ali Javed, Khalid Malik, Aun Irtaza, and Hafiz Malik, observe that these implementations illuminate AI's potential. They note that, “Once you see a machine produce complex functioning code or brilliant images, it’s hard to imagine a future where machines don’t play a fundamental role in how we work and create” (A. Javed, K. Malik, et al. 2020).

More substantive AI implementations are being incorporated deep within existing media production ecosystems. The application Canva Premium includes a model that uses cloud-based AI to remove backgrounds from full motion video scenes (Canva 2023) (*figure 48*). Background removal has traditionally been accomplished using one of two approaches. The most common approach is to film the subject against a solid-colored background and then filter out the specific color from the scene’s chrominance channel. The result is any pixel with that color will appear invisible. As the color green is directly opposite the color values found in human skin tones, most systems are calibrated to the Pantone color 354C, often referred to as “Chromakey Green”. Widely used since the 1960’s in television production (for example, the weather forecast on local television), the technology was originally analog (RCA 1958). The shift to digital at the beginning of the 2000’s simplified the greenscreen process significantly, prompting it to become a staple of corporate and social media videos.

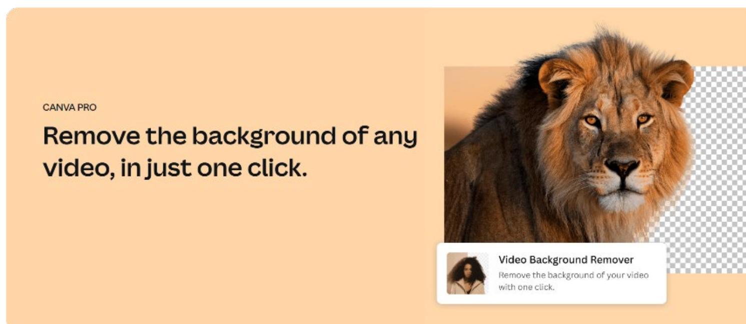


Figure 48 Canva Pro promotional image, Canva website, 2023

The second approach is the painstakingly granular process of tracing a subject’s outline, frame-by-frame, with the purpose of isolating them within a scene. This process of creating traveling mattes, so called because the cutout would travel and change shape as the subject moved in the frame, dates back to the early experimental work of the Lumière brothers. Contemporary applications, like Adobe’s AfterEffects, have automated the process by allowing an artist to identify the subject with an outline, and then using the luminance value of each individual pixel, the software attempts to track the subject as it moves within the frame. Often referred to as rotoscoping, this process of tracing over live-action footage, one frame at a time, is incredibly time consuming. While the accuracy of an automatically generated matte is often close, it nearly always requires a user to step through the scene to “clean up” inaccurate edges.

Canva’s solution is not focused on removing a specific color or tracking how the edges of a subject move through the frame. Instead, they leverage a variation of an AI background subtraction model, called Unscreen, that learns the distribution of pixels in a frame to identify what should be removed (Coldewey 2020). This Dynamic Deep Pixel Distribution Learning (D-DPDL) allows a user to remove the background in any video (Zhao and Basu 2020) (Canva 2023). Pixel mapping represents the capacity of a model to analyze the content of a video file to classify the framing, subject, and movement.

Adobe has been using narrow AI in their marketing and analytic platforms since 2017. Called Adobe Sensei, advanced iterations of the model have been incorporated into media production through the use of speech-to-text models for generating video captions in PremierePro, image modification tools like skin-smoothing, artifact removal and subject identification in Photoshop, and Content-aware filters that allow a user to replace or remove objects in a photo (Helyer 2022). In March 2023 the company unveiled a new AI model, called Firefly (Adobe 2023). This model is currently focused as a generative diffusion model for still images. Using text prompts from within an application like Photoshop, a user has the ability to generate a complete image or to identify elements within the image they would like to replace with generated content.

One feature included in this recent release of Adobe's AI tools is a text editor housed within Adobe's video editing platform, PremierePro (*figure 49*). Powered by the Sensi large language model, Adobe's implementation of text-based editing allows a creator to edit their video using the same logic they would use when editing text in a word processor. The Sensei model uses a speech-to-text engine to transcribe the dialog within a user's editing timeline. The resultant transcription is displayed in the edit interface, where cutting, pasting, or moving passages of text results in corresponding edits in the video story (Philpott 2023). This model of mapping audio via text to specific frames of the video is similar to the approach I am proposing in my storytelling model. As discussed previously, the compute required to process video is significantly more taxing than the compute required to process audio. If a model maps key locations within a video file by associating those locations to specific words, accessing those locations becomes appreciably more efficient during the process of compiling available footage or locating important moments. The keyword association

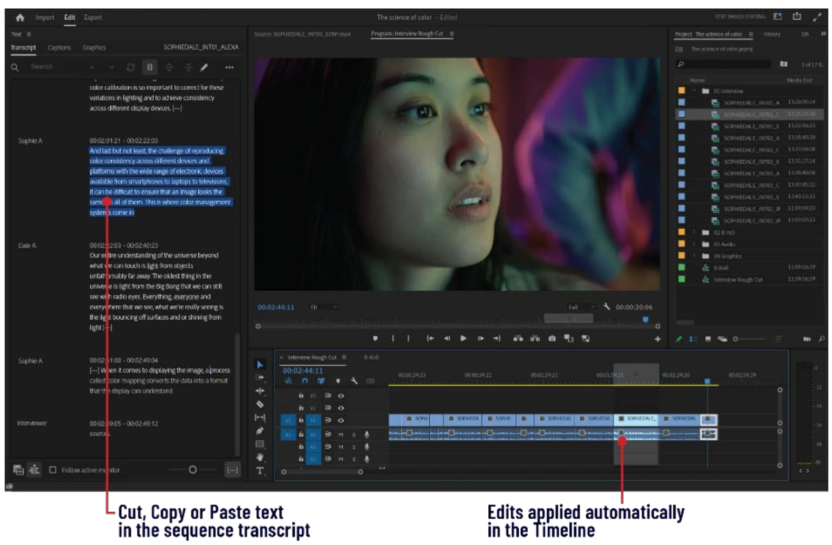


Figure 49 Adobe PremierePro text-based editing interface, Adobe blog, 2023

becomes, in essence, mile markers on the side of the road as the storytelling model nimbly navigates story creation.

Figure 50 Screen shots from Runway video demonstrating prompt-based editing., Runway, 2022

The third application, Runway, has also developed video tools that focus on postproduction challenges. Their generative video model that was discussed in chapter 4, Gen-1, is only one component in an entire suite of AI video tools. Like Canva Pro, they have developed a robust video background removal process, called Green Screen (Germanidis 2020). They have created an AI model that uses a series of neural networks to track objects as they move within the frame (Esser, Michael and Sengupta 2022). And they have developed a model called Soundify that identifies an object in a scene using a zero-shot image classification neural network, retrieves the appropriate sound effect from a database, and matches it to the motion in the frame (Lin, et al. 2021). While all of these applications address specific narrow challenges a creator is likely to face during production, I believe the most significant improvement to the creative workflow is prompt-based editing (*figure 50*). Runway has demonstrated a proof-of-concept application where a user manipulates their footage through written descriptions (Runway 2022). These prompts have the ability to retrieve a specific clip from a user's media database, alter the look of a clip by applying a video effect, and execute a command like 'remove object' to eliminate a distracting element (replacing a sign by cloning surrounding pixels). Because most professional video editors quickly and efficiently operate an application like Resolve or Premiere Pro using their keyboard (commands issued by pressing a key associated with a specific action), they are unlikely to adopt what would be perceived as a less efficient and slower editing process. However, the opportunity to describe an intended outcome in the same way a user would generate an image in Stable

Diffusion is likely to appeal to many who are unfamiliar with the craft of video editing. This approach could be a frictionless entry for those who are overwhelmed by the learning curve.

Runway's prompt-based video editing application has a direct correlation to the proposed Storytelling model, as it represents the capability of an algorithm to execute editing commands, and to process prompts via a GPT engine.

These three examples are but a few of the recent developments that represent important advancements in some of the components required to generate a story via algorithm. Subject identification, prompt-based content generation and text-based editing each leverage functional architectural elements. While the intricacy of the model is rife with dependencies and requires each module to not only perform an assigned task but to then prompt the next task, the rapid pace of iteration demonstrates the potential feasibility.

5.3 Storytelling Model – Proposed training schema

This Storytelling Model builds on the work of many researchers who are making significant leaps forward in AI-based video analysis. Some of the most significant research has taken place in the task of performing content analysis of sports video. While this topic may not initially seem to be relevant to video storytelling, footage from sporting events is a very good framework for training video recognition systems. The camera angles are relatively similar, no matter the venue. The use of replays is consistent, the language used to describe exciting events is based on the common terminology within the sport, and the graphics often appear in similar locations on the screen. For sports broadcasters, there is a pronounced need to automate the process of creating highlight videos. This is particularly true with the sport of cricket, where the average match runs about 7 ½ hours.

It is estimated that over 2.5 billion people follow cricket (Smith 2023). Their interest creates a demand for broadcast coverage. With the number of events covered during a season, broadcasters are seeking to avoid the costs associated with devoting the personnel required to scan 7 ½ hours of footage, searching for a few seconds of action that encapsulate a match. For the past two decades the exploration of approaches for automating sports video analysis has focused on effective solutions for replay detection, shot classification, key-events detection, and summarization (A. Javed, K. M. Malik, et al. 2020). Early work includes research by Kolekar and Sengupta into development of hierarchical classification frameworks (Kolekar and Sengupta 2006), and the application of deep learning to process a complex 5-level analysis of color, subject and edge detection (Midhu and Padmanabhan 2018). Within the last four years three papers have been published that not only address improvements in sports

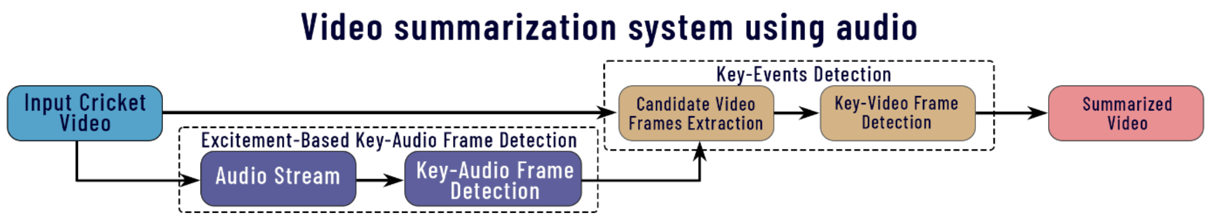


Figure 51 Video summarization system, adapted from
"Multimodal framework based on audio-visual features for summarization of cricket videos", Javeed, et al. 2019

video summarization but represent foundational elements that could be included in the Storytelling Model.

In 2019 Ali Javed and a team of researchers presented a paper that offers what they describe as, "a computationally efficient and an effective sports video summarization framework" (Javed, et al. 2019, 7). Their multimodal framework identifies key events in a video through both audio and visual cues (*figure 51*). By tracking the excitement level registered in the audio signal (of both the commentary from the announcers and crowd reaction), the team identifies potential areas in the video for analysis. In the second stage of analysis, the areas that have been identified are then evaluated to see if the visual elements within the frame match the criteria associated with key events. Each event that matches the criteria is passed to a decision tree classifier, and the video key frames (the beginning and end of each event) are compiled to create the highlight video.

The following year, Javid led a team that refined the model, extending detection beyond audio detection to shot classification (A. Javed, K. M. Malik, et al. 2020). This decision tree framework analyzed the composition of each shot, determining the camera's framing and identifying specific objects in the frame. By extracting the histogram for each shot, a linear display of a specific signal's intensity, they were able to extract signatures for different items (*figure 52*). This digital fingerprint becomes a numeric value that identifies the boundary of an object, with the result being a dataset used for training.

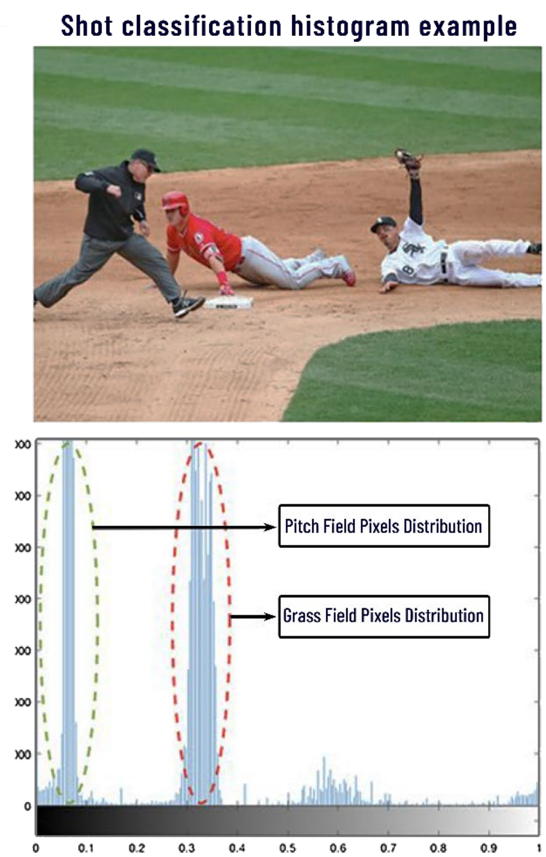


Figure 52 Example of using histogram data from vide to identify objects in the frame.
Adapted from Javed, Malik, et al. 2020

In 2021 Hansa Shingrakhia and Hetal Patel put forward what they described as a hybrid machine learning-based model for cricket video summarization (Shingrakhia and Patel 2021). Their paper acknowledges the computational advantages of using audio for detecting highlights, but they point out this method does not identify semantically meaningful content. Identifying these important events requires knowledge of the specific sport. Their model initially extracts the moments of excitement using an adaptive threshold, speech-to-text framework, and Stacked Gated Recurrent Neural Network with Attention Module (SGRNN-AM). The selected scenes are classified with a Hybrid Rotation Forest Deep Belief Network (HRF-DBN). Using the onscreen graphics indicating the score, the model refines the identification of the characters and action. Finally, this data is fed back into the SGRNN-AM to identify the key moments that will be compiled into the final video. Shingrakhia and Patel point to this hybrid approach resulting in significantly improved accuracy when identifying the most important moments in a match.

Each of these three examples illustrate the capabilities of an algorithm to classify video. They also leverage the smaller computational load of audio processing to locate key moments in video for further analysis. Finally, they blend a sequence of modules, each performing a narrow task, into a coherent whole. Like the three previous examples from testing existing media production tools, these three sports highlights models leverage architectural elements that are important to the functionality of the Storytelling Model.

There is also work that is being done to identify the sentiment of video stories. Deb Roy and Eric Chu have developed an algorithm that analyzes the emotional arcs of movies through the sentiment of the images and sounds (Chu and Roy 2017). In their paper "Audio-Visual Sentiment Analysis for Learning Emotional Arcs in Movies", they work to quantify the emotional impact of stories. Using deep convolutional neural networks, they identified audio and video patterns that were consistently present in scenes that generate a particular emotional response. They accomplished this by initially using human evaluators to annotate video clips for emotional sentiment. This file/text dataset was then used to train the model for sentiment analysis, constantly refining the algorithm by comparing its sentiment rating against the initial human control group. Having a model with the capacity to successfully identify the sentiment of a particular scene, they investigated the 'universal shapes' of emotional arcs in movies and videos, clustering them to identify the distinct classes of stories. Their model demonstrates that an algorithm can be trained to identify the patterns required to successfully tell a story. It can understand the construction of a specific scene and to find the patterns needed to assemble those scenes into a narrative arc. They also believe by understanding what elements are needed for successful story construction, their model can anticipate how an audience will respond. The potential is a model could do more than simply understand how a story is constructed. It could use that information to compile the available elements into what it believes will be the most effectively engaging story.

Disentangling the patterns of story types requires a model that understands each genre, including where particular events occur in a story. That macro task requires ever increasing levels of fidelity through subroutines, models that are constructed on a network of micro tasks. Consider the cascade of dependencies. To generate a story, the model will need to understand the patterns of effective narratives. The consequence of this rather simplified assumption is it will need to deconstruct the order and duration of a myriad of different elements. To deconstruct these elements, it will need to have the ability to identify the different parts. These parts could be the words used in a script, the audio from an interview, the framing of specific shots, the length of each shot, the order of each shot, the order of scenes, the sound of the environment, and the use of music. The complexity of the task will require a multitude of subroutines, each focused on identifying and extracting one thread. Knowing what thread goes where is the ultimate goal for training the model. It will require an algorithm capable of learning from a variety of sources.

Because they are able to use natural language to learn from a wide range of visual concepts, CLIP models are likely to have the flexibility required as the core of the Storytelling Model. In visual tasks, they have the capacity to perform zero-shot analysis on a variety of media. In a study conducted by Radford, Sutsjever, et al., researchers measured CLIP's zero-shot performance on over 30 different datasets. This included tasks such as fine-grained object classification, geo-localization, action recognition in videos, and OCR.B (Radford, Sutskever, et al. 2021). It is even showing the ability to perform optical character recognition on images, meaning it has the potential to extract meaning from screen text the same way it can learn from text strings entered as data.

During training of the proposed Storytelling Model, a source media file is initially broken into two data streams – audio and video. Each of the two streams will be used to extract the different datapoints used for classification. Like many of the highlight video models discussed, the audio stream will be used both as a content source and as a locator for cross-referencing the identification of important segments of video.

The audio present in video stories falls into different categories. One of the predominant audio types is narration. Most often this entails someone reading a prewritten script that describes the action on the screen, provides information, or acts as a narrative transition from one scene to the next. In a news story, much of the information shared with a viewer is presented as the reporter's voice, narration heard over the pictures that comprise the scene. A person talking while on camera, either directly or to another person, is another common category of audio. This can take the form of an interview, a reporter, or an on-screen presenter. While these “talking heads” may appear similar, each perform different functions in a narrative. A reporter is the foundation of a narrative, while an interview is usually a subject. A presenter could be offering instructions on how to perform a task, or it could be a personality sharing events from their day. Another type of audio category is actualizations, audio that appears naturally during the filming and is used to provide additional information. This might be someone talking as they perform a task, or a conversation as someone is walking down the street. Environmental sounds are also an important element to video storytelling. The sound of waves crashing on the beach, the siren of a passing ambulance or kids on a playground can be used to introduce a scene or transition from one idea to another. Finally, music is often used to frame the emotional tone of a story. It affords cues to a viewer how the director expects them to feel as the story unfolds.

To train the model it must first unravel the audio stream into words and sounds. In yet another model designed to extract sports highlights, only this time it is Rugby, Baijal, et al. created a model that filters based on detecting acoustic events (Baijal, et al. 2015). They used a Gaussian mixture model (GMM) to classify the audio into speech and non-speech components. By sifting the stream and extracting only the instances of speech, the model can feed this media into a speech-to-text algorithm (*figure 53*). Or it can use the audio waveforms as sonic fingerprints to identify the type of audio present.

Speech recognition is one of the oldest AI models. The earliest known documented speech recognition system was developed by Bell Laboratories in 1952 (Meng, Zhang and Zhao 2012). By the 1980’s the development of the Hidden Markov Model (HMM) increased system capacity from a few hundred words to several thousand (Rabiner and Juang 1986). Over the past decade, numerous enterprises, such as Apple, Google, and Microsoft, have developed and deployed Automatic Speech Recognition (ASR) systems. These systems function by analyzing an audio stream bit-by-bit, parsing the complex acoustic signal to identify known discrete linguistic units referred to as phonemes. A phoneme is the smallest distinct unit of sound in the human language. The identified phonemes are then

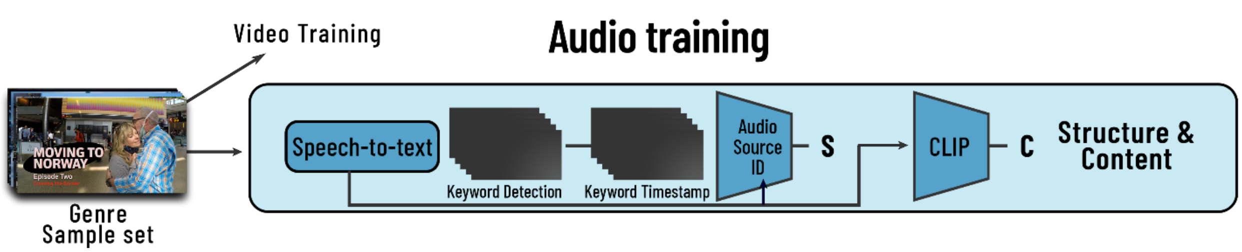


Figure 53 Audio processing flow of Storytelling Model during training

processed through a database of common words, phrases, and sentences as the algorithm uses mathematical models to predict the most likely words or phrases that match the input. Finally, these matches are output as a text stream.

Until recently, AI models have struggled to match human performance (Srinivasan 2022), with the best performing models still misunderstanding one in every twenty words. In 2022 a team from OpenAI labs presented a new model called Whisper (Radford, Kim, et al., Robust Speech Recognition via Large-Scale Weak Supervision 2022). This ASR was trained on 680,000 hours of multilingual and multitask supervised data, with about a third of the dataset's audio featuring non-English speakers. The model uses an encoder-decoder Transformer, where the audio is split into 30-second chunks, converted into a spectrogram, and then passed to an encoder. OpenAI's researchers show that Whisper is able to perform language identification, phrase-level timestamps, multilingual speech transcription, and speech-to-text translation, with an error rate five times lower than previous models (ibid. 2022).

By leveraging an ASR like Whisper, the Storytelling Model framework will afford the algorithm the capacity to detect keywords common within the genre. When a keyword is detected, it can be associated with the timestamp where it appears in the file. A content atlas is created that maps keyword instances to where they can be indexed in the source. The filtered audio can also be processed to determine the audio type, or source identification. Understanding if a particular text string is an interview, a demo, a voice over, or an actuality recorded in the field is an important classifier when learning the patterns of a genre. Ultimately this Audio Source ID will be cross-referenced with the video classifier to help identify video types. Finally, the output from the speech-to-text module is also sent to a CLIP engine for mapping the significance of each element. This information will eventually be output to the Sorting Engine, where it is combined with what has been learned from indexing the video.

While the audio follows one path for analysis, the video follows a separate avenue for classification. To effectively produce a coherent story, an algorithm will need to detect and identify important moments in the source footage (Edwards 2022a). To arm the model with the knowledge required -- for it to know what it is looking for -- the Storytelling Model will need to learn how to recognize image types. The accuracy of an algorithm is defined by the quality of the dataset used during training. These algorithms require meaningful data to improve their accuracy and reduce computational complexities (Shingrakhia and Patel 2021). As we have seen from previous research, the potential for a deep learning framework to automatically process a video stream is prompting the research community to search for the optimum approach for shot classification. In 2015, Raventós, et al. presented an automated framework for creating summaries of soccer matches (Raventós, et al. 2015). If one were to train the Storytelling Model on stories that are an unbroken sequence of shots, there is a risk of cluttering the knowledge base with inconsistent input. The model would have a difficult time finding coherent patterns in the data. Raventós, et al. leveraged a model called shot boundary detection to identify when a scene cuts from one camera angle to another. By identifying these cut points, they were able to break their video stream into smaller segments. Segmenting a video significantly increases the fidelity of the training data, allowing the process of video analysis to be an apples-to-apples comparison for the model. Javed, Malik, et al. indicated this process is, “an indispensable requirement to precisely detect the key-events” (Javed, Malik, et al. 2020, 7242).

In the Storytelling Model, after the training video source file is input into the algorithm for analysis, it will be segmented using Clip Boundary Detection. The beginning and end timestamps for each clip will be compiled into a database before the segments are classified (*see Figure 54*). There are multiple potential solutions available for determining scene and shot classification, and it is likely that the model will leverage one or a combination of approaches. As previously discussed in the analysis of research into cricket video highlights, the model could reference luminance values, graphics placement, and image contrast. A clip can be analyzed for the color distribution using a histogram, like Pei and Chen's approach for indexing tennis and baseball videos (Pei and Chen 2003). Bain et al. presented a learning model that treats clips as 'frozen' snapshots of video (Bain, et al. 2021). This approach significantly reduces the computing load in the early stages of deep learning. By learning to categorize the visual elements contained within a single still image, with the goal of then applying that knowledge to full clip, an algorithm can acquire classification skills in less time. There are also commercial multimedia applications in the market, including Intel's Deep Learning Streamer Pipeline, an open-source cloud-based framework used to detect, classify, track, identify, and count objects, events, and

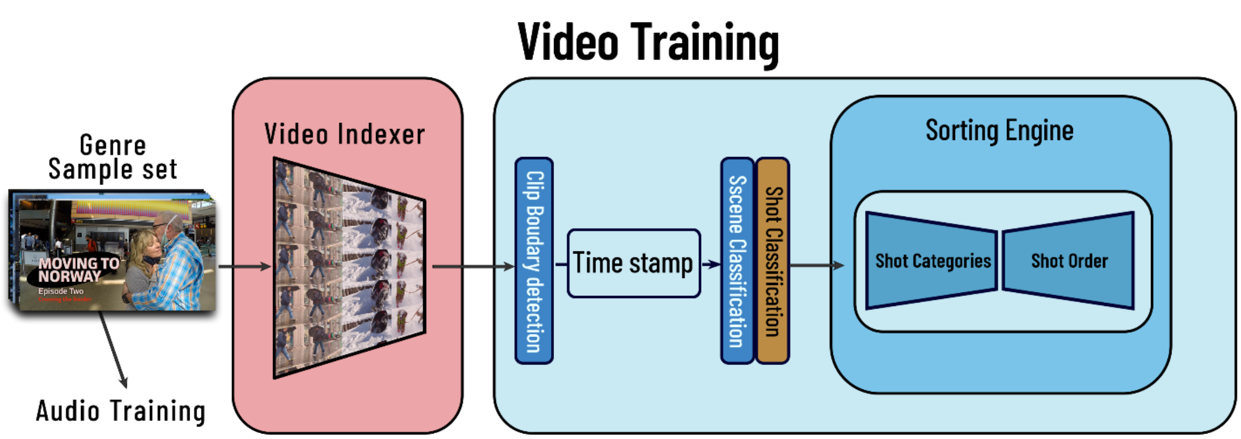


Figure 54 Video processing flow of Storytelling Model during training

people in video streams (Intel 2022). Microsoft offers computer vision analysis designed to optimize manufacturing and track inventory. Their Custom Vision is an AI model with the ability to identify and categorize objects, individuals, and environments (Microsoft 2023).

Although none of the models I have described are plug-and-play solutions, they do demonstrate that developing an algorithm for scene identification, shot composition recognition, and subject identification does not require inventing entirely new technology. Rather, it is an iterative task that builds upon existing research. The rapidly evolving wealth of options illustrates that the process of reading and comprehending the elements bounded by a video's frame is well within reach.

The aim of bifurcating this audio and video stream is to efficiently train the model to identify and classify the common components of video storytelling. However, merely possessing a glossary of keywords, shot types, and durations will not make this proposed model a storyteller. The model will need to utilize what it has learned in the first phase of training to unravel the patterns of storytelling. By having an atlas of words and images, the model will then possess the knowledge to classify the patterns used in scene construction. Equipped with the keywords and phrases commonly found within a genre, it will be capable of classifying an unknown video source, and subsequently learn the patterns of story construction. The model will understand the expected narrative grammar and syntax of effective storytelling.

There are inherent challenges with understanding the nuances of story structure. Like all the creative activities we have explored, storytelling is an art form. There is a precedence for using Artificial Intelligence for automated story generation. These models are most often applied in computer game mastering, where AI manages a player's experience via an interactive narrative (Hernandez, Bulitko and Hilaire 2014). Using an AI experience manager, these models dynamically shape the story during gameplay. Because the algorithm references a prompt defining the player's emotional destination, these models will computationally predict a player's response as it synthesizes the story. In a paper describing "Emotion-Based Interactive Storytelling with Artificial Intelligence" (ibid. 2014), authors Sergio Poo Hernandez, Vadim Bulitko, and Emilie St. Hilaire, write that "stories that are consistent with both the player's actions and authorial goals can be automatically generated via AI planners" (ibid. 2016, 146)

While precedence for AI storytelling exists, developing a model that fabricates a video story is a substantial leap from managing game mechanics that draw on a prewritten database of character responses. A significant challenge will be training the model to understand a story's sentiment. One approach for solving this problem was developed by Damian Borth, Rongrong Ji, Tao Chen, Thomas Breuel, and Shih-Fu Chang. They applied adjective-noun pairs as descriptors for image classification,

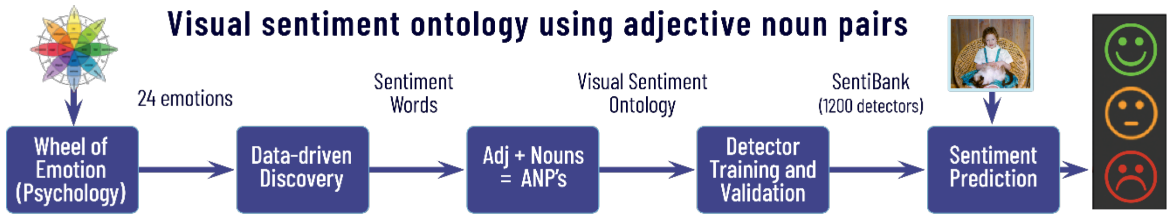


Figure 55 Adapted from "Large-scale visual sentiment ontology and detectors using adjective noun pairs" (Borth, et al. 2013)

leveraging this dataset as the ontology for training a model that discerns image sentiment (Borth, et al. 2013). The researchers used a method built on psychological theories to construct a large-scale Visual Sentiment Ontology (VSO). With this dataset of Adjective Noun Pairs (AN-P), they designed a visual concept detector library called SentiBank to successfully categorize input media (*figure 55*). In the Chu and Roy paper, they used the SentiBank dataset for classifying their training media. They approached the task by training their system on two different types of video stories: a corpus of Hollywood films and a collection of user-generated content from the web (Chu and Roy 2017). Both of these datasets have clearly defined storytelling arcs displaying common filmmaking techniques to convey plot and elicit emotional responses. Using both long-form and short-form stories afforded them a path for training their model to find commonalities in story structure, regardless of duration.

The model will need to comprehend common sentiments linked to different image categories and how the combination of narration and images generate a specific emotional response. By understanding how sentiment, shot composition, and narrative pacing combine to engage a viewer, it will acquire some of the rules it will ultimately apply in the second model.

When generating a story, a significant factor in shaping each narrative arc will be the utilization of a large language model. To achieve this, the model must first be equipped with a search criterion. The story generation prompt will rely on applying the trained model's knowledge (such as genre, composition, audio, sentiment, etc.) to the raw footage ingested by the user, and applying the criteria developed during training to index and rate the media. The complex deep learning process is critical because the data produced from this analysis, combined with the genre guidelines provided by the CLIP engine and the user's initial prompt, will facilitate the creation of a story construct. The instructions utilized by CLIP during generation will also include a sentiment analysis of the source footage.

The computational gymnastics involved in compiling this type of generative narrative may initially seem beyond the capabilities of current large language models like GPT. However, recent developments in Contrastive Reinforcement Learning are bringing the capacity to tell complex stories to AI. Most automated story generation models are limited by the constraints of preferences trained into the algorithm. As a consequence, users often have to engage in personal prompt engineering to generate their intended outcome. In a recent study titled "Robust Preference Learning for Storytelling via Contrastive Reinforcement Learning," the authors presented a training approach designed to align the stories generated by the algorithm with narratives that have received positive critiques by humans (Castricato, et al. 2022). By rewarding the model for generating content that mirrors high-quality narratives preferred by humans, they have developed a robust system of story generation. Incorporating this type of bi-encoder algorithm into the Storytelling Model would increase the likelihood that it will have the capacity to manage the complexities of blending text with video in the second half of the model – story generation.

5.4 Storytelling Model – Proposed story generation schema

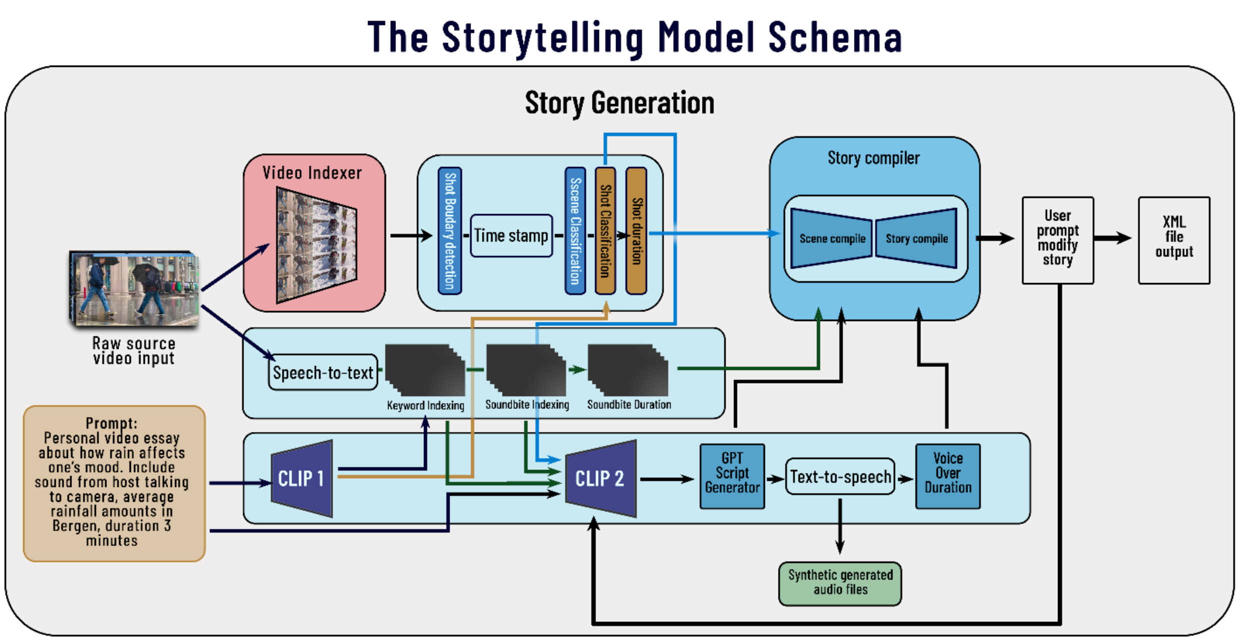


Figure 56 Schema for proposed generative Storytelling Model

The co-founder of ClipChamp, a consumer video editing application, recently highlighted the need for a model that can apply knowledge of story construction to raw footage. Anna Ji stated that users of their application often struggle to create content and don't know where to start when editing a video (Ji 2022). She believes that users will become more successful as video storytellers when they are presented with a recommended way to combine their footage. Ji notes that "There is a need for this, especially in education and gaming" (ibid.).

As an instructor and consultant, I have observed that many potential content creators face a common problem - they lack the interest or time to develop the skills required for effective video storytelling. By applying AI and machine learning to media production, it may be possible to achieve the goal of telling a good story that matches viewer expectations. The question at hand is how we can combine these technologies to read and understand source footage and then weave selected video clips into a coherent narrative.

The architecture of the proposed generative Storytelling Model is complex, reflecting the numerous decisions a creator must make as their story evolves. These decisions are shaped by various factors, including story goals, available assets, and genre tone. Moreover, each choice is informed by what has come before in the story, ultimately determining what will come after. Crafting a video story involves a thousand little decisions that contribute to the whole. The model shares a trait that is common to many algorithms in that it seeks to emulate the human creative process. Given the numerous dependencies, parallel actions, and feedback loops involved in the model, it is crucial to isolate some of the steps for clarity.

The process begins with two inputs. One is a prompt entered by the user defining the genre, tone, length, and context for their intended story. The second is ingesting the source footage for analysis.

When a user inputs their footage, each file is identified by filename and separated into individual video and audio files (*figure 57*). The two signals are separated as they each provide different datapoints. Because a video file is comprised of a sequence of images playing at a set cadence, and most have continuous timecode embedded in the stream, and because the audio file will have exactly the same duration and runs at a consistent sample rate, both the newly created files will have matching timestamps. This data will prove critical later when referencing precise locations in the original video file. The video signal will be indexed using the file's metadata, including filename, duration, format, framerate, and creation date. The audio will also be indexed for name and duration before the clip begins processing. It is at this stage where moments of human speech are identified and segmented. The segments of dialog are then sent to a speech-to-text model for transcription.

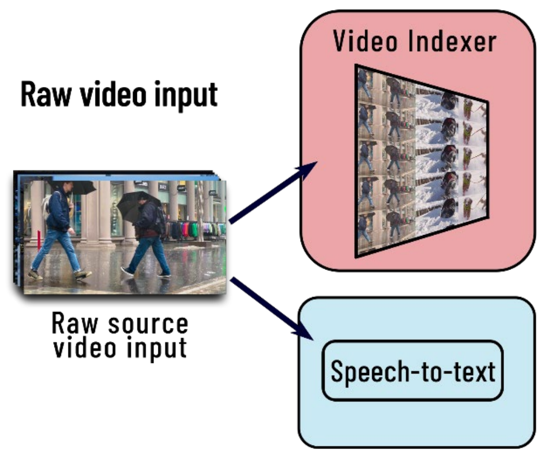


Figure 57 Media ingest path

Once the media is ingested by the engine, the user will enter a prompt describing the intended outcome of their story (*figure 58* Like all generative models, the level of detail describing mood, audience, outcome, genre, video duration, and structural elements like scene identification or defining the video's reporter/host will improve the outcome. One potential approach for assisting the user would be the incorporation of a user interface that prompts them to specify their story's intended emotional or informational moments in their prompt. This approach of identifying concrete narrative events has been successfully incorporated into video game AI story management (Hernandez, Bulitko and Hilaire 2014). There is an opportunity to significantly improve the fidelity of the output by providing the user with a clear way to identify their goals, potentially including where in the narrative arc they intend to see a specific event.

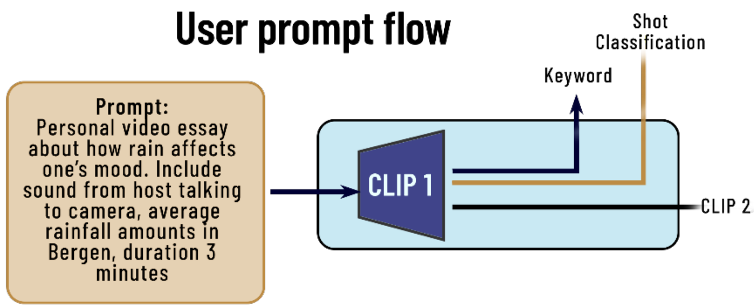


Figure 58 User prompt entry and flow

The prompt will feed into the first CLIP engine where keywords and image tokens will be matched to the deep learning knowledge base. These keywords will be sent to the audio processing engine, where they will be used as guides for identifying key moments in the audio. The video tokens will be sent to the video classification system for use in classifying, identifying, and mapping the available visual components. The initial user prompt is also sent to the second CLIP engine for later use in story generation.

Armed with the keywords from the first CLIP engine and the dialog file segments, the audio processing engine converts the dialog to text, mapping each identified word to the original source timestamp (*figure 59).* The resulting text file is then indexed for relevant keywords and phrases based on the user's original prompt. Segments that contain matching keywords are identified and analyzed as a whole, with the goal of identifying contiguous strings of speech that convey relevant ideas, emotions, and dialog. Each soundbite is assigned a weight based on its alignment to the desired keywords. If the content closely aligns it will receive a high rating, with less alignment resulting in a lower weight. The rated soundbite information is then sent to the second CLIP engine for incorporation into the prompt. An atlas is created using the beginning and end timestamps for each potential soundbite segment, indicating where the corresponding audio is located in the original file. This information is stored for eventual cross-checking against the processed video.

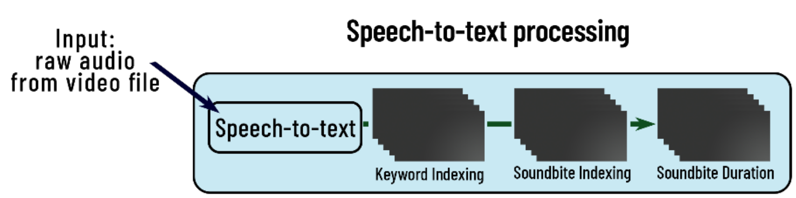


Figure 59 Speech-to-text and soundbite processing

As the audio files are processed, the video files are analyzed separately and in parallel (*figure 60*). Like the audio stream, the indexed video file is segmented, and a Shot Boundary model is used to identify high-fidelity moments in the stream. The initial file is then broken into subclips for processing, with beginning and end timestamps for each segment compiled into a database. The clip then enters an identification phase.

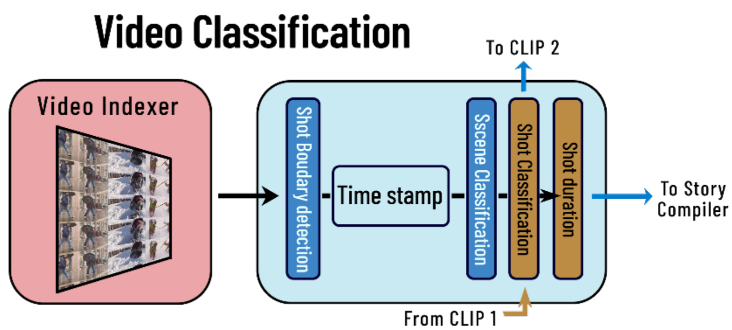


Figure 60 Video indexing and shot classification

During the identification phase, the first model analyzes the footage based on image sentiment, rating the clip according to the model's knowledge of genre and emotion. Next, a second algorithm identifies the framing of the shot, any objects within the frame, and rates the clip using criteria sent from the first CLIP engine (driven by the original user prompt). The results from this analysis are then sent to the second CLIP engine, to be incorporated into the story generation prompt. Finally, the duration of the clip is added to the video database. As a result of this process, each clip now has datapoints that indicate the original file name, the beginning and end of the segment, its sentiment classification, shot description, rating according to the initial prompt, and the clip's duration.

With this compilation of available video and audio assets, the model has the data required for script generation (*figure 61*). A new prompt is created for the second CLIP engine using the original user prompt, a list of audio keywords, a list of soundbites with relevancy ratings, and a list of relevant shots. Before incorporating a soundbite into the narrative, the model compares the rated video with the rated audio, giving more weight to the soundbite with higher rating coincidence and increasing the likelihood of its inclusion in the script. The generated script includes written narration and dialog captured on location and has timestamps indicating the filename and location for each clip incorporated into the story.

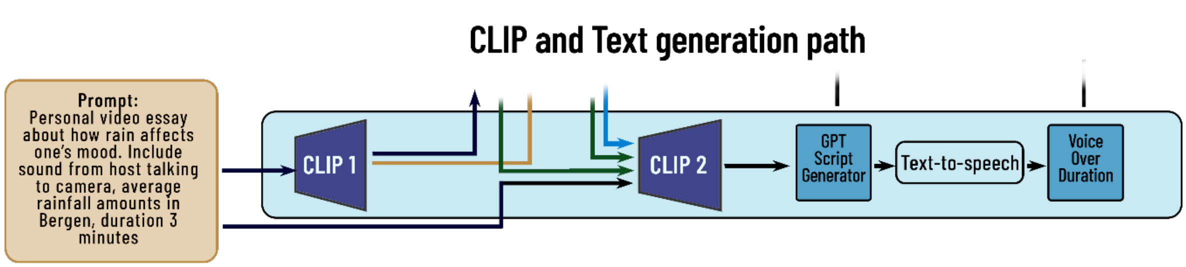


Figure 61 GPT story generation phase

The premise of story generation based on an AI model understanding narrative structure from existing media is not unique. As discussed, developers have created algorithms with the capacity to understand and apply narrative arcs in computer games. There are even commercial services on the market that use AI to help storytellers shape their stories with suggestions.

One such service is Dramatica, a story engine that generates unique plot structures based on a writer's prompt (Write Brothers, Inc. 2022). On their website, they claim that their AI model "is the only writer's tool that can tell you things about your story you didn't tell it" (ibid. 2022). They state that they offer 30,000 individual forms of story structure and provide insights to the writer on how to create an original story. It is worth noting that Dramatica's AI implementation is not a generative model that churns out a complete story. Instead, it creates a basic story structure of plot points, character development and tone, leaving the writing to the human author. This is one approach that could be leveraged by the Storytelling Model to develop the overall arc of the video before the individual scenes are created.

Once the model has generated a script, the narration elements (individual paragraphs) are labeled and sent to a text-to-speech generator to produce segments of synthetic narration. The generated audio files are named according to the assigned script label, with that information and duration data sent to the story compiler. The files are then saved for retrieval by the user.

What is produced by the CLIP and Text generation path can best be described as a wrappers around the metadata of filenames and locations of all source media, the filenames and duration of synthetic narration, and the story divided into distinct scenes. This information is passed on to the Story Compiler for the final phase (*figure 62*).

The Story Compiler assembles the video's components, using input from four primary sources: the dataset of available video assets from the Video Classification engine, the location of audio and soundbites from the Soundbite Duration processor, the duration and filenames of the synthetic audio from the Voice Over module, and the script information from

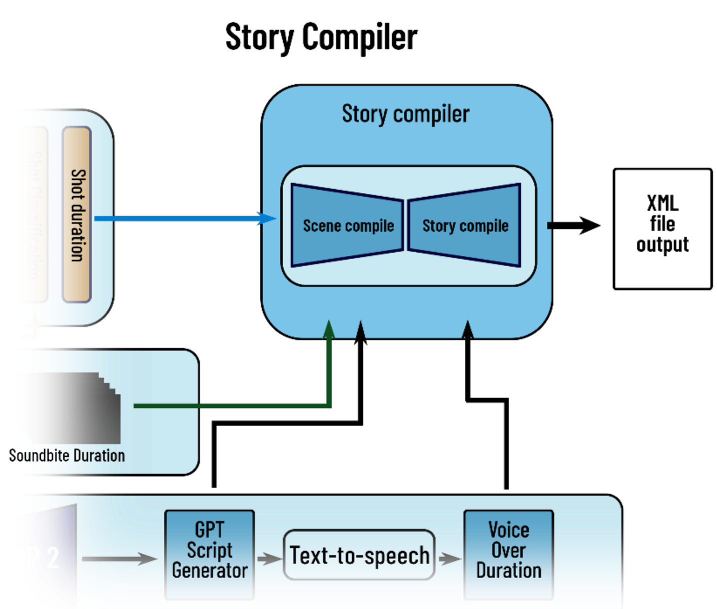


Figure 62 Story compiler

the GPT language model. This scripting information will act as the scaffolding for compiling all assets into the suggested video. To accomplish the task, the compiler will use the script as a framework for breaking the story into modules. These modules will eventually be assembled into a complete video.

The first task is to identify all video clips that have been incorporated into the script as soundbites and mark them in the database as unavailable. This prevents the model from inadvertently repeating an instance of media in the story. To begin the process of assembling the story, the algorithm uses the scene identification from the script to select a scene and gather the required audio assets (synthetic voice-over and soundbites). The compiler determines the total length of the scene by totaling the duration of the synthetic audio and soundbite files, as this will dictate how much video will be needed to illustrate the scene. Using this duration, the model determines what footage is best suited for the scene. To make this determination, it applies its genre and sentiment knowledge to evaluate each potential clip in the database, measuring the anticipated need against the ratings generated in the classification phase. Applying its understanding of shot order and shot length dictated by genre, the compiler stack-ranks the available files and selects moments within the best clips to create a sequence of shots. The clips that have been incorporated into the story are marked as "used", removing them from the database of available footage.

This process of compiling the necessary resources and determining the order of elements for an individual scene is a self-contained action. As previously mentioned, video storytelling relies on small content modules that fit together to create the whole. Each scene connects to the next like building blocks. The data that identifies and describes individual clips exist as packets that can be ranked, ordered, and organized as needed. Although the footage may have been ingested as linear events, the representation of the original files becomes tokens, descriptive packets of data that can be randomly accessed from the atlas. Once the Story Compiler model has identified which scenes need to be assembled, the fabrication of each scene module can be performed in any order. The model could be instructed to prioritize the fabrication of one type of scene before others, thereby ensuring that the best footage will appear in that scene regardless of its placement in the narrative. Alternatively, it could be instructed to assemble the scene sequentially, from beginning to end. The process of assembling the story involves the model working its way through the generated script scene-by-scene.

When the scenes have been compiled, the algorithm sequentially assembles them into a completed story. The output produced by the Storytelling Model is not a video stream, but a sequential table of data that contains the filename and beginning and end points for each piece of media used to create the story. This Edit Decision List (EDL) can be thought of as a lightweight recipe for how to combine the elements, minus the ingredients. While the model uses the original media for indexing, for story generation it processes only the metadata to reduce the compute required. A critical component to the success of this type of architecture is protecting the integrity of the original file name and the timestamps for the media included in the EDL. These two data points afford a media playback or editing platform a consistent path for locating and displaying the correct media. They are similar to geolocation, where the filename and timecode are the media's longitude and latitude.

The generated story EDL has two potential paths. It can be wrapped into an XML file format for use in most video editing platforms, or it can be sent to an interface that displays the script (generated text and selected soundbites) for user review.

Text-based video editing is an emerging mechanism in production applications. As discussed earlier, Adobe's Premiere Pro is integrating this feature directly into its platform (*figure 49*). Gling, a standalone application, also uses AI to analyze the audio track of source footage and provide suggestions on what can be eliminated (gling 2023). With gling, the audio signal is sent through an algorithm that analyzes the audio spectrum to identify moments of silence or interference, noting the timestamps of those instances. The audio is also passed through a speech-to-text engine to generate a transcript of the dialogue. The model evaluates the result of this analysis and selects what it believes are the best "takes" from the recording. This information is presented to the user in an interface where the text is displayed sequentially, with segments that have not been chosen presented as light gray text (*figure 63*). A timeline at the bottom of the interface displays the video segments, and the edited video plays in a window. With this simple interface, a user can select or deselect words in the text display, adding or removing video from their timeline. To cater to a growing global user base, the application currently supports multiple languages, including English, Spanish, Portuguese, French, Russian, Turkish, and Hebrew. Gling has been developed with an output path similar to the proposed Storytelling Model, with the final edited string exported as an XML file.

Gling text-based editing interface

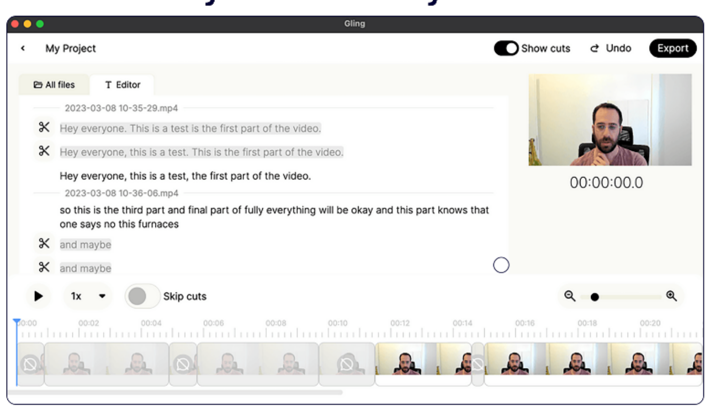


Figure 63 Gling text-based editing user interface.

To ensure the effectiveness of the proposed Storytelling Model, it is essential to include a feedback loop that enables users to modify their original prompts. As identified at the beginning of this paper, the premise of this model is to generate a *suggested* video story that serves as a starting point for creators, providing agency to inject their own point of view and refine the story during both the generative and editorial phases. Furthermore, as described in Chapter 2, the unpredictable nature of generative algorithms is the catalyst driving the emergence of "prompt engineering." When writing their initial prompt, a user has specific expectations about the outcome. Unfortunately, often the algorithm's interpretation of their word choice produces unexpected results, and it will need to be revised.

It is likely that the initial story generated by this algorithm will be unsatisfactory, and the prompt will need to be modified. The creator can choose to enable a user interface where they will be presented with their original prompt and given the opportunity to augment their text. Additionally, they will be presented with a text display of two different elements. The first is a transcript of their source media with the status of the media clearly identified (used or unused in the story). Here, they can specify elements to include or ignore. The second is the generated script. Once again, they will have an opportunity to adjust the language or reorder elements within the story. These changes will be incorporated into a modified prompt that replaces the initial user prompt and is sent to the CLIP 2 engine (*figure 64*). CLIP 2 will have stored the compilation of video and audio assets generated through video, audio, and soundbite classification, allowing the model to repeat the process. It will generate a new script using GPT, fabricate the synthesized audio, compile the story, and display the result.

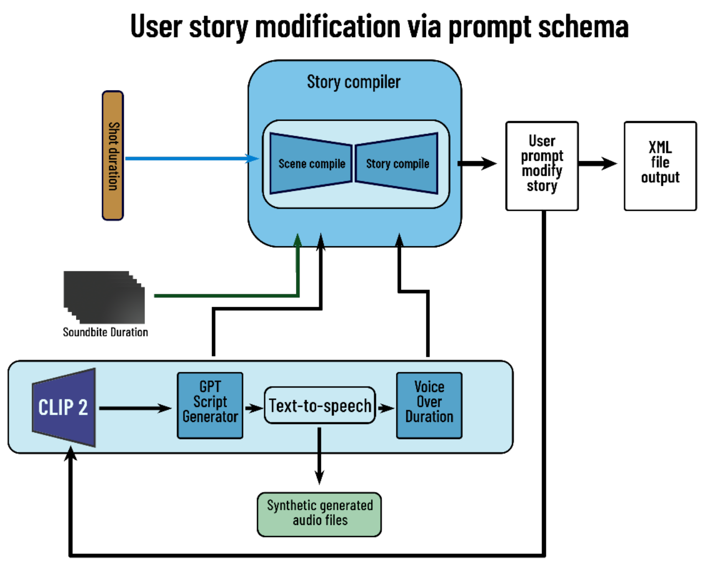


Figure 64 Story modification

When the creator is satisfied with the suggested story, the final result will be output as an XML file for import into a third-party editing application.

5.5 Storytelling Model – Going forward

As stated at the beginning of this chapter, this model is based on a multitude of assumptions. Fundamentally, many of the core technologies required as components are either in development or available in the market. However, due to the model's complexity, there is a near certainty that most of these parts will be initially unable to communicate with each other. In previous media production and application testing projects, I have learned that what appears to be a simple task on the surface often breaks the build. Therefore, this proposal is not focused on presenting a detailed blueprint of the model's connective tissue. Instead, it presents a broad architectural view of how an algorithm could generate a suggested video story. It does not address the thousands of details, assumptions, or technological leaps that are required for a near term realization. Rather, it attempts to model the inherently human choices made by video editors and their operational behaviors when crafting a story. These choices are shaped by their personal experience, training, and exposure to thousands of video stories. They learn what works, and what does not, by watching and doing. Training the model to understand successful narrative arcs, common features of genres, and how framing and sequencing shape a visual narrative is an effort to emulate the knowledge that informs each human creator's editorial choices.

I anticipate that the costs associated with development and deployment could be prohibitive, given that video analysis technology is still either limited or expensive. However, recent developments suggest that this field is rapidly evolving, which could substantially reduce the cost of required computation. We are also seeing significant advancements in graphic hardware capabilities. It is likely that the latest releases of GPUs can perform the tasks of ingest and analysis on a local computer. Processing the video and audio analysis locally, with only the resulting data sent to a neural network, will reduce the required upload bandwidth, network storage, and overall computational load. These savings in time and resources could drastically decrease the cost of deployment.

I firmly believe that the development of a video storytelling model is imminent. While my proposed architecture is one viable approach, many others are almost certainly sketched out on whiteboards and cocktail napkins. As a species, digital video is a testament to the fact that we are inherently driven to tell stories. The positive outcome of Artificial Intelligence could be the potential to enable those with creative ideas but lack the required skills to bring their imaginations to life. These tools could open creative avenues for those storytellers who have neither the opportunity nor interest to be a film student and lack the time to master a complex and nuanced craft. I believe that presenting users with suggestions that kick-start their creative process will empower a new group of storytellers. AI has the potential to unlock abundance from scarcity, and this abundance can be measured by human creativity. An AI video storytelling model has the potential to reveal an abundance of stories. Stories previously only imagined.

Chapter 6

Conclusion

When we pose the question of whether Artificial Intelligence is creative, it likely refers to an inquiry into whether AI can create like us. The history of AI's development repeatedly reveals proponents who are attempting to fabricate machines that mirror human intelligence and creativity. The symbiotic relationship between cognitive studies and advancements in deep learning is intractably linked. Even the architecture of these machines is a massive model of synaptic activity. However, creative computing is more than a mirror of neuroscience. It represents a shift in how we present our ideas, our feelings, our reflections of self, and its impact will forever change the archival breadcrumbs we leave behind. Going forward, creative products are less likely be the unique effort of human hands and imagination. Instead, they will be a hybridized amalgamation comprised of how a creator sees their world blended with the ideas and experiences of millions of others captured in a training dataset.

6.1 Findings

I began this journey by asking, "Is it possible to create an AI model that provides a streamlined path for video production to a storyteller?" Could a user input source footage and frame their intended story outcome in a prompt, expecting the algorithm to analyze the available media and apply its deep learning to generate a recommended story? Essentially, are the foundational parts available to begin the process of creating this model? I believe most of the core elements currently exist. However, our realization of a video storytelling model is on an indeterminate time horizon, its arrival framed by iterative refinements to the parts, the inherent difficulty in connecting the pieces, and, like most technological leaps, the capital required for development. While this idea responds to a need, it is uncertain whether there will be enough of a market to justify the leap. A massive model with this sort of complexity could easily fall victim to the outcome of a corporate cost-benefit analysis, an analysis that may very well be performed by AI.

While the realization of my proposed algorithmic Storytelling Model may not be a near-term event, I believe the questions posed during my exploration are relevant to contemporary manifestations of synthetic creativity. From painting, to graphics, to prose, to music, to full-motion video, the capacity of AI models to generate work comparable to humans is approaching indistinguishable levels. As humans, we display our creativity through the process of combining associative elements into a new form (Chapter 1.3), a behavior that is also the foundation of generative AI. While the creativity expressed in generative work does not leverage an individual artist's lens of human experience, for most viewers, the message that is delivered resonates nonetheless. Research performed by Payel Das and Lav Varshney, Mark Coeckelbergh, and Joo-Wha Hong and Nathaniel Curran (1.4) shows that the judgment of quality and aesthetic value with AI-created work is comparable to human-created work, and for most viewers, the artistic source is essentially irrelevant. It appears then, through this determination, that AI can create. However, these creative abilities differ from how humans create. Our appreciation of AI's creative expression reflects how their aesthetic and emotional messages resonate with our perceived projections of intelligence.

AI is ushering in a broad societal transformation. Its development has the potential to mirror the unchecked deployment of social media platforms, posing consequences that are equally detrimental. The technology's adoption rate, by both individuals and corporations alike, is rapidly accelerating (1.8). Today's embrace of AI is driven by both novelty (among users) and the potential profits to be gained from market share (by corporations).

Generative AI's unprecedented scale of deployment is likely to act as a catalyst for the loss of creative jobs (1.9). The prevailing belief is that when a machine can accomplish the same task as an artist, musician, or writer, and reduce costs to near zero, it becomes impossible for human creatives to compete in the marketplace. As a result, we are likely to see more creative workers shift their focus from internal creative exploration to the task of developing techniques that can harness AI's benefits, while differentiating their products from the avalanche of derivative AI-generated output. For those who continue in their roles as creative workers, the challenge will be sustaining a unique creative voice.

The incorporation of Artificial Intelligence into media production is not theoretical; these tools have already been integrated into production paths (2.1). With the rapid advances in capability, creatives are beginning to ask whether generative AI is focused less on helping artists and more on replacing them (2.2).

Creative workers' concerns over lost relevance and diminished market value are a refrain that has echoed with every industrial revolution. With each leap forward specific occupational sectors have seen job loss tied to the introduction of automation. When mechanical looms were introduced to handmade lace production in 1809, it devastated a cottage industry that at its peak had employed over 100,000 artisans (2.3). The rise of steam-powered factories saw manufacturers realize significant reductions to their production cost. This rise in efficiency reduced the demand for labor, with the number of workers displaced by automation growing in lockstep with profits (2.4). It took over fifty years before the demand for labor recovered, and for workers to see the value of their real wages rise. This pause was repeated with each technological leap. The Age of Science and Mass Production saw the invention of airplanes, electric power, and the assembly line. Again, job creation was centralized in factories, with most located in urban areas. Small communities were devastated as the population flocked to the cities. Skilled craftspeople, most located in rural areas, were pushed out of the markets, unable to compete with the low-cost/high-volume merchandise spilling out of factories (2.5).

The Digital Revolution saw the creation of jobs requiring specific technical skills. Assembly line workers were replaced by automated systems, and office clerks saw their numbers dwindle as one employee was expected to do the work of ten (2.5). From 1987 to 2017, worker displacement dramatically outpaced reinstatement. The jobs destroyed by the shift to digital, most of them requiring manual labor skills, would never return. The new jobs created were high-skill jobs requiring qualifications that many workers were unable to access.

Each of these industrial revolutions resulted in improved labor efficiency, resulting in a decline in workforce demand. All repeated the cycle of job loss with a decades-long pause before job recovery was realized. Despite recent press releases where corporations are espousing a belief that the AI coming revolution will empower employees and make them more efficient (2.9), if the past is prologue, then we are likely to experience a significant global job loss due to Artificial Intelligence (2.7). Creatives will certainly feel this impact, particularly with the potential of generative Artificial Intelligence to create at scale with minimal cost. To survive as artists, creatives will need to adapt to the changing landscape. It is almost certain they will need to find ways to incorporate these tools as avenues for both ideation and efficiency.

These tools of artificial intelligence are rooted in mathematical formulas of logical reasoning (3.1). This true/false framework provides a systematic representation of human cognition, using on/off switches modeled on the human neuron (3.2). The term "artificial intelligence" encompasses a broad technological universe, with subcategories that define the capabilities of different models. The AI we encounter daily, such as home voice assistants, recommendations in our streaming media feed, and generated images on the web, is narrowly focused on performing specific tasks (3.1). This narrow AI is being broadly deployed with much fanfare as stand-alone applications and quietly woven into existing software as features. One important challenge with most of these implementations is that researchers are finding it difficult to explain an algorithm's behavior. The shift to deep neural networks (DNN) has come at the cost of sacrificing our understanding of exactly how they work. Like the inner workings of the human brain, developers can only guess at what the discrete parts of the digital brain are doing. These networks are designed to mimic the processes that form human intelligence. However, they also seem to mirror human behaviors, such as opaque thinking and inexplicability. Addressing problems like bias, where the training data's racial and gender bias results in the same biases woven into the fabric of the AI model, becomes difficult to fix when developers do not understand how a system works.

Contemporary efforts to construct a “thinking machine” were sparked when neuroscientist Warren McCulloch and logician Walter Pitts released their model of the human neuron in 1943. The race to create a working theoretical model that would describe a functional cognitive process progressed in fits and starts for the next 80 years. Scientists and researchers, including pioneers like Alan Turing, W. Ross Ashby, John McCarthy, Marvin Minsky, and others, struggled to develop algorithms that could produce demonstrations of intelligence within the constraints of the existing hardware's limitations (3.4). Turing favored a symbolic approach, with a string of symbols entered into the computer, processed, and the algorithm outputting a resultant string of symbols. Others, like American psychologist Frank Rosenblatt, favored the more "connectionist" neuron model, where learning was facilitated through the creation and modification of connections between simulated neurons. Until the 1980s, the symbolic approach held favor, but researchers determined that the capabilities of these models were inherently limited. By 2012, neural modeling had become the dominant approach, principally because it had the potential to apply more computational cycles to researchers' algorithms. More cycles meant more math, which meant more robust algorithms. This capacity to throw more math at these complex computational problems has led to a string of recent advancements in Artificial

Intelligence. Developers have gained deeper understandings of natural language models, natural language generation, "Few-Shot" and "Zero-Shot" deep learning models, and diffusion models (3.5).

For-profit global companies are funding much of this development, lured by the potential of incorporating a slew AI features into products, with the expectation that rolling Artificial Intelligence into every corner of their applications will differentiate their applications from the competition (3.6). With this corporate rush to scale, economists have begun predicting the displacement of millions of workers worldwide. The World Economic Forum anticipates that by 2028, 85 million workers will have been displaced by Artificial Intelligence. This race to position AI into a broad spectrum of markets, with global organizations pushing the technology ever forward, and workers realizing that their livelihoods are in peril, is a clear indication that users have little say about when, and if, AI is woven into their lives.

Artificial Intelligence is beginning to roll through the creative community like a tidal wave. The capacity to generate synthetic sound, images, and video has the potential to disrupt the representation of creativity, personal expression, and the ways in which artists reflect cultural values to ourselves (4.1). The technology promises to free us from the mundane so that we can focus on the complex. However, it also portends a time when current modes of creative expression are devalued. Perhaps the most widely talked-about implementation of generative AI, and the models with the widest market deployment, are text generators. These algorithms process vast amounts of text in search of patterns, with the goal of applying the knowledge gained to generate statistically probable outputs. The text output from these generators presents a human-like presence through the appearance of intention and thought. But that appearance is an illusion (4.1). They lack the capacity to understand or verify the source of their training data. They do not have the capacity to learn the definition, meaning, or etymology of words, nor do they understand the associations between the words used and their corresponding referents. Generated texts appear to present ideas with authority, but the veneer of intellect is only illusory, stemming from the model's fundamental inability to process meaning. The output from large language models is little more than a predictive string of words, not ideas. Humans use language to communicate; we write to convey ideas, we choose specific words to convey emotion, and we craft word order to reflect the rhythm of our voice. The shortcomings of these models shortchange the experience of presenting original ideas.

This intellectual shortcoming also extends to generated visual art. The capacity to create images, painting, illustration, photography, and drawing has been a demonstration of human intelligence for centuries. The art of painting has been described as understanding the spectrum of artistic styles, with an artist possessing the knowledge to express their creativity through a specific technical approach. For the last ten years, Generative Adversarial Networks (GAN) have been the standard approach for generating images with artificial intelligence (4.1). Unlike a natural language model, which learns the probability of word order to predict the next one through a discriminative process, generative algorithms predict the features that are required to generate a label. Discriminative algorithms are about *finding* patterns, whereas generative algorithms are about *creating* patterns.

The process of training a GAN is inherently difficult. The generated images do not always imitate life, particularly when attempting to render lifelike human features (4.2). When refining a generated image, the model may choose a path of creation that it believes is correct. Unfortunately, because of an inherent randomness hard-wired into the model, the final output may reveal the refinement path was incorrect, resulting in mangled and misaligned human forms. This refining inaccuracy can result in asymmetrical eyes, circles drawn as ovals, and rather comical renderings of human faces and hands. An additional problem includes the utilization of copyrighted images and artwork when training these AI models, raising concerns over potential legal infringement on the original creator's rights.

An alternative approach for generating synthetic images is OpenAI's Contrastive Language–Image Pre-training, or CLIP (4.3). This model blends diffusion with OpenAI's deep understanding of Natural Language Processing. Their system architecture, called DALL·E, incorporates techniques from machine learning, computer vision, natural language processing (NLP), natural language generation (NLG), and synthesis. A primary benefit of CLIP is that it does not require significant levels of human supervision, thus reducing preprocessing of the training dataset. Instead, it learns visual concepts via supervision by their GPT natural language model. By mapping the intersection of image and language, it affords CLIP the ability to correlate relationships, acquiring computational associations of aesthetics, style, and a broad knowledge of descriptive language.

The developments in text and image generation described above are foundational elements required for using AI to process, augment, and even generate video. Of particular note, video models that have incorporated the capacity to analyze a video clip's content and structure have led to algorithms that can describe a source video's appearance, semantics, and identify specific objects in a scene (4.4). The integration of CLIP, natural language, and video classification models has a direct bearing on the potential to create a Video Storytelling algorithm.

Most people find that video creation is difficult. Crafting a coherent narrative can feel opaque as video production requires a functional literacy in storytelling, directing and editing (5.1). A creator needs to understand the architecture of storytelling, identifying their story's expected outcome and defining what will be required for the story's beginning, middle and end. What many creators don't grasp is that most genres of online video follow specific patterns, patterns that can be recognized by AI. Recent advancements in AI processing and analysis of video point to the possibility of designing a storytelling model. The model would leverage machine learning to recognize patterns in a specific genre of video stories, such as patterns of story structure, narrative content, sentiment, and image sequencing. This knowledge will be applied to raw footage input by a user, with the goal of outputting instructions for creating a video that matches the structure of the chosen story type. The model follows a two-phase approach, Training phase and Generation phase.

Many technical elements used in the proposed model currently exist in commercially available video production tools (5.2). Subject identification, prompt-based content generation, and text-based editing all leverage functional architectural processes required for training and generation.

Research into content analysis of sporting event video recordings points to a number of functional processes leveraged in the algorithm. This includes the ability to use text as a guide for locating critical moments in a video, using text to identify key words, the ability to classify objects in the frame, the ability to discern scene or camera changes, the capacity to generate a compressed video summary of the analyzed source material, and the wisdom of applying decision trees for choosing specific moments for inclusion in the summary video (5.3). CLIP models, as described in Chapter 4, are an effective approach for applying Natural Language Processing to image analysis, genre identification, sentiment analysis, clip identification, and image categorization. Natural Language Processing allows for the synthesis of user prompts, keyword application, and script generation (5.4). Speech-to-text models will enable the model to extract dialogue from source footage for processing by the NLP. Synthetic audio generators will create narration from the generated script. Large-scale Visual Sentiment Ontology libraries, such as SentiBank, represent an avenue for classifying genre sentiment training videos, image sentiment of training videos, and classification of the user's media ingested as the source for their story. Much of the foundational technology required for a Storytelling Model exists as commercially viable applications or proof of concept research.

While the proposed model is predicated on several contingencies, the value of this video storytelling proposal may be less in near-term prototyping and more in identifying a starting point for developing the algorithm. The complexity of connecting these many discrete functions into a working whole would be daunting. A likely path would be to take a modular approach, developing individual components with an eye towards how the output will feed into the next module. Like the architecture of the model itself, the development approach would be executing autonomous subroutines, refining these small parts, eventually assembling them into the whole.

It is also important to acknowledge that the costs associated with development and deployment of the algorithm could be prohibitive. To justify the time and effort it would be wise to pursue user studies to identify potential market size, use cases, and revenue potential. As has been displayed throughout the history of computing, the costs associated with hardware, network requirement, compute times, and coding, will be reduced over time. I believe as the cost of development continues to fall, we will realize the deployment of AI video storytelling assistants within the next 5 years.

6.2 Effort heuristic

The goal of this thesis has been to explore the viability of a video storytelling model. This journey has carried me through applications of cognitive theory and explorations of what it means to be creative. The likely parallels of job loss between contemporary creatives and the craftspeople of history are clear. Disentangling AI's development illuminates that the iterative progress of technology is never linear. Instead, the fits and starts are framed by overcoming hardware limitations, persistence, and finding solutions to complex problems with unconventional thinking. Each path traveled on this rather incongruent journey has led me to believe in the potential practicality of the Storytelling Model.

The recent advancement in Artificial Intelligence also leads me to conclude we are reaching the knee in the curve of development. Ray Kurzweil described the knee as a tipping point, the time where society notices technology's rapid change (Kurzweil 2001, 38). Awareness spawns interest, which spawns investment, which accelerates even more development. It is the point where the consequences of an ever-accelerating rate of change creates dynamic pressure on society, prompting unintended outcomes – good, bad, or both. Kurzweil's interpretation of the curve, his emphasis on the importance of identifying a specific inflection point, has been criticized as being rather loose with the facts (Modis 2006). Scientists and researchers have described his work as lacking rigor, precise definitions, identified uncertainties, and meticulous data collection (Ayres 2006).

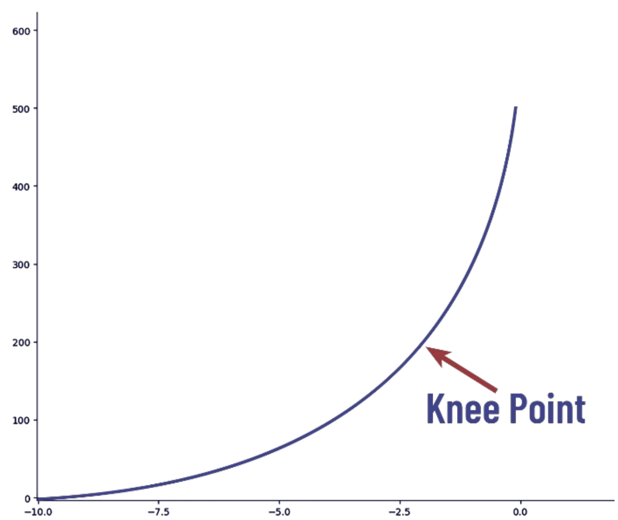


Figure 65 The knee point in an exponential curve

While Kurzweil's work may lack specificiity, his model of change is a useful way to describe our potential future with AI. In times of change, we search for those who have their eyes on the horizon, experts who offer forecasts about the consequential choices we make today. I am reluctant to offer my own conjecture about the future, as I have learned to be skeptical of expert forecasts. When looking back I have found too often these prognostications, be it about technology, economics, or even the weather, are at best only somewhat accurate. In Chapter Two, we saw how periods of rapid change are followed by periods where nothing much happens, or at least not at the exponential rate reflected in the curve above. Technological advances, be it steam power, mass production, or AI, do not operate like a living breathing organism. Technology is not a self-generating force progressing by its own internal dynamics. It cannot be mapped like bacteria in a Petri dish. Even Moore's Law eventually sputtered to a stop.

Despite these inherent limitations, forecasts can be useful as we organize our thoughts about contemporary events. The cycles of history, disruptions in creativity, employment, society, and technology, emerge as patterns. These repetitions can illuminate future outcomes. For this thesis I have explored the iterations of development as a means to frame a potential future. This exploration of where we have been has brought me to the conclusion that our trajectory forward may be filled with dramatic and sudden changes.

Economists assure us that the impact on jobs from AI’s creative destruction will act as it always has, new occupations will be created to replace the old. This scenario is likely to be true in time, but in the near term, this progress will come at a cost. That cost will be paid not only in the form of lost paychecks, but in the personal identity of those effected (Burley and Eisikovits 2022). As discussed in Chapter three, it appears many AI developers give little thought about the occupational impact of their efforts. Creatives caught in the wake of development all too often appear as little more than collateral damage. As a director, photographer, writer, and editor, the practice of the craft of storytelling affords me a means by which I can exercise creativity, judgment, synthesis, and tenacity. I have observed these same behaviors in hundreds of my peers, and I feel I can safely assume that it is a common criterion for determining self-worth among most creatives. Losing these sources of self-identity are likely to damage the individual, yet the impact will be difficult to measure. My desire to explore the potential and peril of AI is not simply an academic interest. It is self-preservation. If the consequences of AI are as dire as I fear, my hope is that research is conducted into identity loss and the constriction of creativity. This path of research could be a means of illumination, identifying the impact of AI-driven job loss on artists.

Those who survive the changes are most likely to find success by discovering ways to insert their humanity into their efforts. Elaine Freedgood notes that when the production of machine-made lace became widespread, customers mourned the loss of the handmade item, ascribing greater value from the perspective of its loss (Freedgood 2003). For the consumer, searching out evidence of the artist's creative mark conveys a preference in the competition between hand-made and factory, quality versus mass-production, where "the dear tries to trump the cheap in arguments that continue to appeal to us not least of all because they seem to offer a concrete resistance to commodity culture, a resistance accomplished by hands and the handmade" (ibid. 2003, 629). There is a term for this human differentiation — the handprints principle (Roose 2022). Proponents believe that there is a direct relationship between evidence of human participation and perceived value. The more obvious the human effort, the higher its perceived worth.

Social scientists refer to this phenomenon as the "effort heuristic." In his book *The Power of Human* Adam Waytz describes the substantial transformational value when physical or digital artifacts are marked by human contact (Waytz 2019). He cites numerous studies indicating a consumer preference for goods and experiences with obvious evidence of human effort. This is true even when the goods and experiences are identical. Yann LeCun writes there is a belief that society will attach more value to goods that exhibit intervention by a human, and less value to goods that are built by robots (LeCun 2017). Kevin Roose notes that effort heuristic "explains a lot about the rise of craft breweries, farm-to-table restaurants, and artisanal Etsy shops" (Roose 2022, 124). Effort heuristic also offers one explanation why mass-produced items, products with no humanizing touch, no evidence that they were crafted rather than punched out of a die-cast mold, are devalued. Writer Frank Bruni describes the differentiation between machine-created and human-created as a means to distinguish our efforts. "The human endeavor to create is the font and province of originality. It's the cornerstone of identity" (Bruni 2022). The devaluation of effort lends a plausible reason why flooding a market with AI-generated artwork represents such a significant threat to artists.

The risks of AI extend beyond a creative's sense of self-worth. We are likely to see further centralization of technology with fewer players. AI's development requires significant investment at scale, and as a result, Big Tech will almost certainly become Really Big Tech. In the near term, we will likely be required to trust corporate claims of good-faith self-regulation. As discussed in Chapter three, many global companies are working to foster internal cultures of responsible AI. Despite these good intentions, history tells us that there will inevitably be pressure on these companies to be the first to market with a fancy new feature, potentially resulting in unimagined consequences for a swath of unsuspecting people. History tells us that Really Big Tech will acquire start-ups, competitors, and tech incubators, becoming Really, Really Big Tech. As public concerns over monopolization and centralized power grow, governments will eventually become involved in regulating the industry, just as they did with the American railroads in the 1880s and Standard Oil in 1906. Governments will eventually regulate AI because the current model of unchecked development points to the realization that no society is equipped to manage the impact that AI will have on citizens and culture.

The rate of advancement with algorithms like Large Language Models is likely to plateau within the next two to three years. With tighter controls on training data, limitations on the hardware's computational capability, and algorithms operating at functional capacity, the inherent limitations with current training techniques will slow progress. Instead, we are likely to experience the infiltration of AI into our daily tasks. Tools like ChatGPT aren't going anywhere; and barring some major regulatory intervention, this particular form of machine intelligence will become a fixture of our society. By automating the mundane, it will become a ubiquitous presence in how we work, how we create, and how we record the seminal moments in life's passage. Perhaps the greater risk comes from the unregulated distribution of some of the source code that powers Artificial Intelligence. While it is unlikely a global corporation would willingly develop and implement malicious AI, there is a legitimate fear that malevolent actors or rogue states will seed disruptive media and malicious code into global society. The most disconcerting aspect of this scenario is we may not even know it is there.

6.3 Close

Creation is never fixed; it is dynamic and subjective. If we look at the evolution of art, new technologies have consistently altered how we express our insights, emotions, and experiences. They have challenged and enriched artistic viewpoints, changing how we create. Photography, for example, altered portraiture; filmmaking altered photography; video altered filmmaking; streaming altered video; and now, AI is changing the foundation of creative expression.

Painters were threatened by the fidelity and immediacy of photography. There were concerns that photography would invade people's privacy because, unlike sitting for a painted portrait, a photographer could capture someone's likeness without their knowledge. There were even legal arguments over the copyright of photographs and who could distribute and reproduce images (Rogers 1883). The case Sarony v. Burrow-Giles Lithographic Co. established the right of a photographer to copyright their work, in the same way that artists in other mediums could protect their work. It legally established photography as a legitimate form of artistic expression.

Just like Photoshop, a tool that enabled the digital manipulation of photos and artwork, thus creating new work derived from other sources, AI is likely to become little more than another resource used by visual artists. The provenance of the training data will be known, the creator will have ownership, and copyright will be assigned. What differentiates AI's generative creativity from previous leaps is that these models are ever-changing, never static, and always learning, attributing the creator may prove to be far more difficult.

This thesis is a record of a particular point in time as new developments, new research, and new complications are presented daily. I believe the world of storytelling that I have known for the last five decades is about to be turned upside down. Both the opportunities and challenges with AI creativity are immense—not just for creators but for society. How will it redefine our perceptions of storytelling? How will it redefine how we reflect our experiences, intellect, and emotions? When artist Gokul Pillai used Midjourney to imagine seven of the world's wealthiest people living in the slums of India, one can argue his work transcended from a digital parlor trick to social commentary (*figure 66*) (Pillai 2023). One can contend that his commentary is perhaps a little too "on the nose," relying on stereotypical perceptions of poverty. Still, for me, the irony that these men, many of whom have invested heavily in the development of artificial intelligence, are placed in an alternate universe of poverty, and created with AI, leads me to see their entitlement in a different light. These images have inherent humanity, yet they were not created by human hands. These images tell a story. Which brings us back to where we began. Are they a demonstration of creativity? Is this art?

Figure 66 Slumdog Millionaires, Gokul Pillai, Instagram posts, April 7, 2023

Works Cited

Abram, Cleo. 2022. "The REAL fight over AI art." *Huge If True.* New York, New Yotk: YouTube, September 14. Accessed September 19, 2022. https://www.youtube.com/watch?v=NiJeB2NJy1A.

Acemoglu, Daron, and Pascual Restrepo. 2020. "The wrong kind of AI? Artificial intelligence and the future of labour demand." *Cambridge Journal of Regions, Economy and Society* 13 (1): 25–35. Accessed December 10, 2022. doi:10.1093/cjres/rsz022.

Adobe. 2023. "Adobe Unveils Firefly, a Family of new Creative Generative AI." *Adobe.* March 21. Accessed May 20, 2023. https://news.adobe.com/news/news-details/2023/Adobe-Unveils-Firefly-a-Family-of-new-Creative-Generative-AI/default.aspx.

—. 2023. *Power incredible experiences with AI.* Accessed February 4, 2024. https://www.adobe.com/sensei.html.

Agarwal, Sakshi, and Lav R. Varshney. 2019. *Limits of Deepfake Detection: A Robust Estimation Viewpoint.* Research findings, arXiv. Accessed December 2022. doi:10.48550/arXiv.1905.03493.

Akyürek, Ekin, Dale Schuurmans, Jacob Andreas, Tengyu Ma, and Denny Zhou. 2022b. *What Learning Algorithm is In-contex Learning? Investigations with Linear Models.* Research findings, arXiv:2211.15661v2, 1-28. doi:10.48550/arXiv.2211.15661.

Alammar, Jay. 2016. "A Visual and Interactive Guide to the Basics of Neural Networks." *GitHub.* December 14. Accessed March 3, 2023. http://jalammar.github.io/visual-interactive-guide-basics-neural-networks/.

Allen, Robert C. 2009. "Engels' pause: Technical change, capital accumulation, and inequality in the british industrial revolution." *Explorations in Economic History* (Elsevier) 46 (4). doi:10.1016/j.eeh.2009.04.004.

Alsheibani, Sulaiman, Yen Cheung Monash, and Chris Messom. 2018. "Artificial Intelligence Adoption: AI-readiness at Firm-Level." *PACIS 2018 Proceedings.* Pacific Asia Conference on Information Systems (PACIS). 37-67. Accessed December 2022. https://aisel.aisnet.org/pacis2018/37.

Andersen, Sarah, interview by Casey Newton and Kevin Roose. 2023. *Elon's Crumbling Empire and Generative AI Goes to Court* Hard Fork. New York. January 20. https://www.nytimes.com/2023/01/20/podcasts/20hard-fork-elon-empire-artists-ai.html.

Anderson, Martin. 2022. "Preventing 'Hallucination' in GPT-3 and Other Complex Language Models." *Unite.ai.* December 9. Accessed January 6, 2023. https://www.unite.ai/preventing-hallucination-in-gpt-3-and-other-complex-language-models/.

Angluin, Dana, and Carl H. Smith. 1983. "Inductive Inference: Theory and Methods." *ACM Computing Surveys* 15 (3): 237–269. doi:10.1145/356914.356918.

Arriagada, Leonardo, and Gabriela Arriagada-Bruneau. 2022. "AI's Role in Creative Processes: A Functionalist Approach." Edited by Fabio Fossa, Caterina Moruzzi and Mario Verdicchio. *Odradek. Studies in Philosophy of Literature, Aesthetics, and New Media Theories* (Università di Pisa) 8 (1): 77-110. Accessed December 2022. https://research.rug.nl/files/234999832/AI_s_Role_in_Creative_Processes_Arriagada_2022_.pdf.

ArtStation. 2023. *ArtStation.* February 9. Accessed February 9, 2023. https://www.artstation.com/.

—. 2023. "Use of AI Software on ArtStation." *ArtStation.* January 20. Accessed February 9, 2023. https://help.artstation.com/hc/en-us/articles/11451085663501-Use-of-AI-Software-on-ArtStation.

Autor, David H. 2014. "Skills, education, and the rise of earnings inequality among the "other 99 percent"." *Science* 344 (6186): 843-851. https://www.science.org/doi/pdf/10.1126/science.1251868.

Ayres, Robert U. 2006. "The Singularity Myth." *Technological Forecasting and Social Change* 73 (2): 95-127. doi:10.1016/j.techfore.2005.12.003.

Baijal, Anant, Jaeyoun Cho, Woojung Lee, and and Byeong-Seob Ko. 2015. "Sports highlights generation based on acoustic events detection: a rugby case study." *2015 IEEE International Conference on Consumer Electronics (ICCE).* Las Vegas, USA: IEEE. 20–23. Accessed December 2022. doi:10.1109/ICCE.2015.7066303.

Bain, Max, Arsha Nagrani, Gül Varol, and Andrew Zisserman. 2021. *Frozen in Time: A Joint Video and Image Encoder for End-to-End Retrieval.* Research findings, arXiv:2104.00650, 1-15. doi:10.48550/arXiv.2104.00650.

Baio, Andy. 2022. "Exploring 12 Million of the 2.3 Billion Images Used to Train Stable Diffusion's Image Generator." *Waxy.* August 30. Accessed February 10, 2023. https://waxy.org/2022/08/exploring-12-million-of-the-images-used-to-train-stable-diffusions-image-generator/.

—. 2022. "Invasive Diffusion: How one unwilling illustrator found herself turned into an AI model." *Waxy.* November 1. Accessed February 4, 2023. https://waxy.org/2022/11/invasive-diffusion-how-one-unwilling-illustrator-found-herself-turned-into-an-ai-model/.

Ball, Willietta Goddard. 1922. "Lace in Portraiture." *The Brooklyn Museum Quarterly* (Brooklyn Museum Stable) 9 (3): 123-133. https://www.jstor.org/stable/26463138.

Barabba, Vincent. 2011. *The Decision Loom: A Design for Interactive Decision-making in Organizations.* Dorset, UK: Triarchy Press.

Bar-Tal, Omer, Dolev Ofri-Amar, Rafail Fridman, Yoni Kasten, and Tali Dekel. 2022. *Text2LIVE: Text-Driven Layered Image and Video Editing.* Research findings, ArXiv:2204.02491, 1-21. doi:10.48550/arXiv.2204.02491.

Bashiri, M., and A. Farshbaf Geranmayeh. 2011. "Tuning the parameters of an artificial neural network using central composite design and genetic algorithm." *Scientia Iranica* 18 (6): 1600-1608. doi:10.1016/j.scient.2011.08.031.

Bawa, A. S., and K. R. Anilakumar. 2013. "Genetically modified foods: safety, risks and public concerns—a review." *Journal of Food Science and Technology* 50: 1035–1046. Accessed 2023. doi:10.1007/s13197-012-0899-1.

Beauchene, Vinciane, and Molly Cunningham. 2020. *The End of Management as We Know It.* Survey results, Boston: Boston Consulting Group, 15. https://www.bcg.com/publications/2020/end-management-as-we-know-it.

Beer, David. 2013. *Popular Culture and New Media: The Politics of Circulation.* New York: Palgrave MacMillan. doi:10.1057/9781137270061.

Belsky, Scott. 2022. "Bringing the next wave of Artificial Intelligence to Creative Cloud." *Adobe Blog.* October 18. Accessed Janaury 24, 2023. https://blog.adobe.com/en/publish/2022/10/18/bringing-next-wave-ai-creative-cloud.

Benjamin. 2016. *Sunspring.* Directed by Oscar Sharp. Produced by End Cue.

Benjamin, Walter. 2006. "The Work of Art in the Age of Mechanical Reproduction." Chap. 3 in *Media and Cultural Studies: KeyWorks*, by Meenakshi Gigi Durham and Douglas M. Kellner, 18-40. Malden, MA: Blackwell Publishing Ltd. Accessed December 2022. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315668505-19/work-art-age-mechanical-reproduction-walter-benjamin.

Berlyne, Daniel E. 1970. "Novelty, complexity, and hedonic value." *Perception & Psychophysics* 8: 279–286. doi:10.3758/BF03212593.

Berman, Ruth A., and Dan Isaac Slobin. 1994. "Narrative structure 1." In *Relating Events in Narrative A Crosslinguistic Developmental Study*, by Ruth A. Berman and Dan Isaac Slobin, 35-81. Hove, East Sussex, United Kingdom: Psychology Press.

Berriman, Richard, and John Hawksworth. 2017. *Will robots steal our jobs? The potential impact of automation on the UK and other major economies.* Economic forecast, PricewaterhouseCoopers, 30-47. Accessed December 2022. https://www.pwc.co.uk/economic-services/ukeo/pwcukeo-section-4-automation-march-2017-v2.pdf.

Biersdorfer, J. D. 2022. "Use That Everyday A.I. in Your Pocket." *The New Yorl Times.* June 29. Accessed January 4, 2023. https://www.nytimes.com/2022/06/29/technology/personaltech/use-that-everyday-ai-in-your-pocket.html.

Böcking, Saskia. 2008. "Suspension of Disbelief." In *The International Encyclopedia of Communication*. John Wiley & Sons, Ltd. doi:10.1002/9781405186407.wbiecs121.

Boden, Margaret A. 2011. *Creativity and Art:Three Roads to Surprise.* Oxford: Oxford University Press. Accessed November 2022.

—. 2004. *The Creative Mind: Myths and Mechanisms.* Second. London: Routledge. Accessed October 2022. https://www.tribuneschoolchd.com/uploads/tms/files/1595167242-the-creative-mind-pdfdrive-com-.pdf.

Boole, George. 1854. *An investigation of the laws of thought: on which are founded the mathematical theories of logic and probabilities.* Walton and Maberly. https://www.gutenberg.org/files/15114/15114-pdf.pdf.

Borth, Damian, Rongrong Ji, Tao Chen, Thomas Breuel, and ShihFu Chang. 2013. "Large-scale visual sentiment ontology and detectors using adjective noun pairs." *Proceedings of the 21st ACM international conference on Multimedia.* Barcelona, Spain: Association for Computing Machinery. 223–232. Accessed 2022. doi:10.1145/2502081.2502282.

Bostrom, Nick, and Eliezer Yudkowsky. 2014. "The ethics of artificial intelligence." Chap. 15 in *The Cambridge Handbook of Artificial Intelligence*, edited by Keith Frankish and William M. Ramsey, 316–334. Cambridge: Cambridge University Press. Accessed December 2022. http://faculty.smcm.edu/acjamieson/s13/artificialintelligence.pdf.

Bovard, Rachel. 2022. "What Makes Trump Different From DeSantis and Other Republicans." *New York Times.* November 22. Accessed November 23, 2022. https://www.nytimes.com/2022/11/22/opinion/trump-desantis-republicans.html.

Boyd, Brian. 2009. *On the origin of stories: Evolution, Cognition and Fiction.* Cambridge, MA: Harvard University Press.

Braidotti, Rosi. 2019. "A Theoretical Framework for the Critical Posthumanities." *Theory, Culture & Society* (SAGE Journals) 36 (6): 31–61. Accessed August 13, 2021. https://doi-org.pva.uib.no/10.1177/0263276418771486.

Briggs, Joseph, and Devesh Kodnani. 2023. *The Potentially Large Effects of Artificial Intelligence on Economic Growth.* Economics forecast, Goldman Sachs, 1-20. https://www.ansa.it/documents/1680080409454_ert.pdf.

Brooks, David. 2023. "In the Age of AI, Major in Being Human." *The new York Times.* February 2. Accessed February 3, 2023. https://www.nytimes.com/2023/02/02/opinion/ai-human-education.html.

Brooks, Tim, Janne Hellsten, Miika Aittala, Ting-Chun Wang, Timo Aila, Jaakko Lehtinen, Ming-Yu Liu, Alexei A. Efros, and Tero Karras. 2022b. *Generating Long Videos of Dynamic Scenes.* Research findings, arXiv:2206.03429, 1-22. doi:10.48550/arXiv.2206.03429.

Broussard, Meridith. 2018. *Artificial Unintelligence: How Computers Misunderstand the World.* Cambridge, MA: The MIT Press.

Brownlee, Marques. 2022. "The Truth About AI Getting "Creative"." *Marques Brownlee.* New York: YouTube, December 9. Accessed December 10, 2022. The Truth About AI Getting "Creative".

Bruni, Frank. 2022. "Will ChatGPT Make Me Irrelevant?" *The New York Times*, December 15. Accessed December 16, 2022. https://www.nytimes.com/2022/12/15/opinion/chatgpt-artificial-intelligence.html.

Burley, Jake, and Nir Eisikovits. 2022. "Workplace automation and political replacement: a valid analogy?" Edited by John MacIntyre and Larry Medsker. *AI and Ethics* 1-10. doi:10.1007/s43681-022-00245-6.

Burley, Jake, and Nir Eisikovits. 2022. "Workplace automation and political replacement: a valid analogy?" *AI Ethics* 1-10. doi:10.1007/s43681-022-00245-6.

Cameron, James, and Gale Anne Hurd. 1984. *The Terminator.* Directed by James Cameron. Produced by Gale Anne Hurd. Performed by Arnold Schwarzenegger. Orion Pictures.

Canva. 2023. *Canva Pro Features.* Accessed March 3, 2023. https://www.canva.com/pro/features/.

Cargill, Thomas. 2023. "Graphic Designers Are Loving These NEW Tools!" *Satori Graphics.* Khiri Khan: YouTube, Janaury 17. Accessed January 30, 2023. https://youtu.be/eTffPm_e1ko.

Carney, Mark. 2018. "The future of work." *Public Policy Forum.* Toronto: Bank of England. 1-28. https://www.bankofengland.co.uk/-/media/boe/files/speech/2018/slides-for-mark-carney-speech-at-public-policy-forum-toronto.pdf.

Carter, Denise. 2018. "How real is the impact of artificial intelligence? The business information survey 2018." *Business Information Review* 35 (3): 99–115. Accessed December 2022. doi:10.1177/0266382118790150.

Castricato, Louis, Alexander Havrilla, Shahbuland Matiana, Michael Pieler, Anbang Ye, Ian Yang, Spencer Frazier, and Mark Riedl. 2022. *Robust Preference Learning for Storytelling via Contrastive Reinforcement Learning.* Research findings, arXiv:2210.07792v1. doi:10.48550/arXiv.2210.07792.

Ceci, L. 2023. "Hours of video uploaded to YouTube every minute as of February 2020." *Statista.* January 9. Accessed January 9, 2023. https://www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/.

—. 2023. *Number of Lensa AI app downloads worldwide from January 2021 to December 2022 MTD.* January 27. Accessed March 23, 2023. https://www.statista.com/statistics/1350961/lensa-ai-app-downloads-worldwide/.

Cellan-Jones, Rory. 2014. "Stephen Hawking warns artificial intelligence could end mankind." *BBC.* December 2. Accessed Janaury 20, 2023. https://www.bbc.com/news/technology-30290540.

Cetinic, Eva, and James She. 2022. "Understanding and Creating Art with AI: Review and Outlook." *ACM Transactions on Multimedia Computing, Communications, and Applications* 18 (2): 1-22. Accessed December 2022. doi:10.1145/3475799.

Chamberlain, Rebecca, Caitlin Mullin, Bram Scheerlinck, and and Johan Wagemans. 2018. "Putting the art in artificial: Aesthetic responses to computer-generated art." *Psychology of Aesthetics, Creativity, and the Arts* 12 (2): 177–192. Accessed December 2022. doi:10.1037/aca0000136177.

Chayka, Kyle. 2020. "The digital death of collecting." *Kyle Chayka Industries.* September 18. Accessed November 6, 2021. https://kylechayka.substack.com/p/essay-the-digital-death-of-collecting.

Chen, Brian X. 2022. "How to Use ChatGPT and Still Be a Good Person." *The New York Times.* December 21. Accessed December 22, 2022. https://www.nytimes.com/2022/12/21/technology/personaltech/how-to-use-chatgpt-ethically.html.

Cheng, Mingyong. 2022. "The Creativity of Artificial Intelligence in Art." *Proceedings of The 2021 Summit of the International Society for the Study of Information.* IS4SI. 1-5. Accessed December 2022. doi:10.3390/proceedings2022081110.

Cherry, Miriam A. 2019. "The Future Encylopedia of Luddism." In *Economic Science Fictions*, edited by William Davies, 327-340. London: Goldsmiths Press.

Chi, Clifford. 2020. "YouTube Analytics: The 15 Metrics That Actually Matter." *HubSpot.* October 19. Accessed May 18, 2022. https://blog.hubspot.com/marketing/youtube-analytics.

Chomsky, Noam, Ian Roberts, and Jeffrey Watumull. 2023. "The False Promise of ChatGPT." *The New York Times*, March 8: A21.

Chowdhery, Aakanksha, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, et al. 2022. *PaLM: Scaling Language Modeling with Pathways.* Research findings, arXiv:2204.02311. doi:10.48550/arXiv.2204.02311.

Chu, Eric, and Deb Roy. 2017. "Audio-Visual Sentiment Analysis for Learning Emotional Arcs in Movies." *IEEE International Conference on Data Mining.* New Orleans, LA USA: ICDM. 829-834. Accessed February 8, 2022. doi:10.1109/ICDM.2017.100.

Cisco. 2020. "Cisco Annual Internet Report (2018–2023) White Paper." *Cisco.* March 9. https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html.

Clark, Evans. 1928. "March of the Machine Makes Idle Hands." *The New York Times*, February 26: 129. Accessed December 12, 2022. https://timesmachine.nytimes.com/timesmachine/1928/02/26/107090960.html?pageNumber=129.

Coeckelbergh, Mark. 2016. "Can Machines Create Art?" *Pilosopy of Technology* 30: 285–303. doi:10.1007/s13347-016-0231-5.

Cognizant. 2017. "21 Jobs of the Future." *Cognizant.* November. Accessed February 7, 2023. https://www.cognizant.com/content/dam/connectedassets/cognizant-global-marketing/marketing-channels/cognizant-dotcom/en_us/insights/documents/21-jobs-of-the-future-a-guide-to-getting-and-staying-employed-over-the-next-10-years-codex3049.pdf.

Cohen, I. B. 1994. "The use of "bug" in computing." *IEEE Annals of the History of Computing* (IEEE) 16 (2): 54-55. doi:10.1109/85.279235.

Colbrook, Matthew J., Vegard Antun, and Anders C. Hansen. 2022. "The difficulty of computing stable and accurate neural networks: On the barriers of deep learning and Smale's 18th problem." *PNAS* 119 (12): 1-10. doi:10.1073/pnas.2107151119.

Coldewey, Devin. 2020. "Kaleido's Unscreen is dead simple drag-and-drop background removal for video." *TechCrunch.* October 2. Accessed April 20, 2023. https://techcrunch.com/2020/10/02/kaleidos-unscreen-is-dead-simple-drag-and-drop-background-removal-for-video/.

Collingwood, Robin George. 1938. *The Principles of Art.* London: Oxford University Press.

Colom, Roberto, Sherif Karama, Rex E. Jung, and Richard J. Haier. 2010. "Human intelligence and brain networks." *Dialogues in Clinical Neuroscience* 12 (4): 489–501. Accessed January 2023. doi:10.31887/DCNS.2010.12.4/rcolom.

Common Crawl. 2023. *Common Crawl Homepage.* March. Accessed March 13, 2023. https://commoncrawl.org/.

Copeland, B. Jack. 2004. *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life: Plus The Secrets of Enigma.* Oxford: Clarendon Press. doi:10.1093/oso/9780198250791.003.0016.

Corazza, Giovanni Emanuele. 2016. "Potential Originality and Effectiveness: The Dynamic Definition of Creativity." *Creativity Research Journal* 28 (3): 258-267. doi:10.1080/10400419.2016.1195627.

Coward, Cameron. 2017. "AI and the Ghost In the Machine." *Hackaday.* February 6. Accessed Janaury 31, 2023. https://hackaday.com/2017/02/06/ai-and-the-ghost-in-the-machine/.

Crabapple, Molly. 2022. "Beware a world where artists are replaced by robots. It's starting now." *Los Angeles Times.* December 21. Accessed Janaury 16, 2023. https://www.latimes.com/opinion/story/2022-12-21/artificial-intelligence-artists-stability-ai-digital-images.

Crevier, Daniel. 1993. *AI: The Tumultuous Search for Artificial Intelligence.* New York, NY: BasicBooks.

Das, Payel, and Lav R. Varshney. 2022. "Explaining Artificial Intelligence Generation and Creativity: Human interpretability for novel ideas and artifacts." *IEEE Signal Processing Magazine* (IEEE) 1053 (5888/22): 85 - 95. Accessed December 2022. doi:10.1109/MSP.2022.3141365.

Daugherty, Paul R., and H. James Wilson. 2018. *Human + Machine: Reimagining Work in the Age of AI.* Boston, Massachusetts: Harvard Business Review Press. Accessed November 2022.

Demos Helsinki. 2021. *Artificial Intelligence 4.0 First interim report: From start-up to implementation.* Interim report, Steering group of the Artificial Intelligence 4.0 program, Helsinki: Ministry of Economic Affairs and Employment of Finland, 1-79. http://urn.fi/URN:ISBN:978-952-327-643-7.

Devereaux, Bret. 2023. "On ChatGPT – A Collection of Unmitigated Pedantry." *Collections.* February 17. Accessed February 22, 2023. https://acoup.blog/2023/02/17/collections-on-chatgpt/.

D-ID. 2022. *Create Your Own AI Video in 30 Seconds.* September 19. Accessed Janaury 25, 2023. https://www.d-id.com/ai-video-generator/.

Dietschy, Sara. 2023. "14 Ai tools creators should actually use." *Sara Dietschy.* YouTube, January 12. https://youtu.be/03iSjMQ3a1U.

Dilmegan, Cem. 2020. *What is Few-Shot Learning? Methods & Applications in 2023.* Research finding, AI Multiple. https://research.aimultiple.com/few-shot-learning/.

Dozier, Rob. 2019. "This Clothing Line Was Designed by AI." *Vice.* June 3. Accessed December 12, 2022. https://www.vice.com/en/article/vb9pgm/this-clothing-line-was-designed-by-ai.

Earnshaw, Pat. 1994. *Lace Machines and Machine Laces.* Dublin: Gorse Publications.

Edwards, Phil. 2022. "How I'm protecting my job against AI." *Phil Edwards.* Richmond, Virginia: YouTube, October 16. Accessed November 14, 2022. https://www.youtube.com/watch?v=EUmgvsGnw0w.

—. 2022a. "How I'm protecting my job against AI." *Phil Edwards.* Richmond, Virginia: YouTube, October 16. Accessed November 14, 2022. https://www.youtube.com/watch?v=EUmgvsGnw0w.

—. 2022. "I found the perfect metaphor for AI Art." *Phil Edwards.* Richmond, Virginia: YouTube, September 11. Accessed November 14, 2022. https://www.youtube.com/watch?v=CfVHXDty5Pk.

Eisikovits, Nir, and Alec Stubbs. 2023. "ChatGPT, DALL·E 2 and the collapse of the creative process." *The Conversation.* Janaury 12. Accessed Janaury 16, 2023. https://theconversation.com/chatgpt-dall-e-2-and-the-collapse-of-the-creative-process-196461.

Eisikovits, Nir, and Dan Feldman. 2021. "AI and Phronesis." *Moral Philosophy and Politics* 9 (2): 181-199. doi:10.1515/mopp-2021-0026.

Elgammal, Ahmed, interview by Sam Charringron. 2019. "Creative Adversarial Networks for Art Generation with Ahmed Elgammal - TWiML Talk #265." *This Week in Machine Learning and AI.* YouTube, (May 13). https://www.youtube.com/watch?v=IatH1Oi4ajQ.

—. 2018. "What the Art World Is Failing to Grasp about Christie's AI Portrait Coup." *Artsy.* October 29. Accessed December 20, 2022. https://www.artsy.net/article/artsy-editorial-art-failing-grasp-christies-ai-portrait-coup.

Elgammal, Ahmed, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone. 2017. "CAN: Creative Adversarial Networks, Generating "Art" by Learning About Styles and Deviating from Style Norms." *Eighth International Conference on Computational Creativity.* Atlanta: ICCC. 1-22. Accessed December 2022. https://doi.org/10.48550/arXiv.1706.07068.

Elgammal, Ahmed, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone. 2017. *CAN: Creative Adversarial Networks, Generating 'Art' by Learning About Styles and Deviating from Style Norms.* Research findings, arXiv:1706.07068, 1-22. Accessed December 2022. doi:10.48550/arXiv.1706.07068.

Eliaçik, Eray. 2022. "Pros and Cons of AI." *Dataconomy.* April 21. Accessed December 16, 2022. https://dataconomy.com/2022/04/risks-and-benefits-of-artificial-intelligence/.

Elliott, Francesca. 2017. "Weaving Numbers: The jacquard Loom and Early Computing." *Manchester Science and Industry Museum.* October 17. Accessed February 2, 2023. https://blog.scienceandindustrymuseum.org.uk/weaving-numbers-jacquard-loom-early-computing/.

Engels, Frederick. 1892. *The Condition of the Working-Class in England in 1844 with a Preface written in 1892.* Translated by Mrs. F. Kelley Wschnewetzky. London: George Allen & Unwin.

Engels, Friedrich. 2001. "The Condition of the Working Class in England." Chap. 1 in *The Sociology and Politics of Health*, edited by David Banks and Michael Purdy, 6. London: Routledge. Accessed December 2022.

Epstein, Ziv, Sydney Levine, David G.Rand, and Iyad Rahwan. 2020. "Who Gets Credit for AI-Generated Art?" *iScience* 1-10. Accessed December 27, 2022. doi:10.1016/j.isci.2020.101515.

Eschner, Kat. 2016. "In 1913, Henry Ford Introduced the Assembly Line: His Workers Hated It." *Smithsonian.* December 1. Accessed February 1, 2023. https://www.smithsonianmag.com/smart-news/one-hundred-and-three-years-ago-today-henry-ford-introduced-assembly-line-his-workers-hated-it-180961267/.

Esser, Patrick, Johnathan Chiu, Parmida Atighehchian, Jonathan Granskog, and Anastasis Germanidis. 2023. *Structure and Content-Guided Video Synthesis with Diffusion Models.* Research findings, Runway Research, arXiv:2302.03011, 1-26. doi:10.48550/arXiv.2302.03011.

Esser, Patrick, Peter Michael, and Soumyadip Sengupta. 2022. *Towards Unified Keyframe Propagation Models.* Research findings, arXiv:2205.09731, 1-13. doi:10.48550/arXiv.2205.09731.

Farrell, Jeremy. 2007. *Identifying Handmade and Machine Lace.* Edited by The Museum of Costume and Textiles. Nottingham: Renaissance Subject Specialist Network Implementation Grant Programme. 39. Accessed November 14, 2022. https://dressandtextilespecialists.org.uk/wp-content/uploads/2021/08/Identifying-Handmade-Lace-and-Machine-Lace.pdf.

Felkin, William. 1887. *A History of the Machine-wrought Hosiery and Lace Manufactures.* London: Longmans, Green, and Co. Accessed December 2022. https://ia800305.us.archive.org/34/items/ahistorymachine00felkgoog/ahistorymachine00felkgoog.pdf.

Fillmore, Charles J. 1976. "Frame semantics and the nature of language." *Annals of the New York Academy of Sciences: Conference on the origin and development of language and speech* 280 (1): 20-32.

Florida, Richard. 2014. *The Rise of the Creative Class--Revisited: 10th Anniversary Edition--Revised and Expanded.* New York: Basic Books. Accessed November 2022.

Freedgood, Elaine. 2003. "Fine Fingers: Victorian Handmade Lace and Utopian Consumption." *Victorian Studies* (Indiana University Press) 45 (4): 625-647. Accessed November 14, 2022. http://www.jstor.com/stable/3829530.

Frey, Carl Benedikt. 2019. *The Technology Trap: Capital, Labor and Power in the Age of Automation.* Princeton: Princeton University Press. doi:10.2307/j.ctvc77cz1.

Frey, Carl Benedikt, and Michael A. Osborne. 2017. "The Future of Employment: How Susceptible Are Jobs to Computerisation?" Edited by Oxford Martin Programme on Technology and Employment. *Technological Forecasting and Social Change* 114: 254-280. Accessed December 12, 2022. doi:10.1016/j.techfore.2016.08.019.

Fuegi, John, and Jo Francis. 2003. "Lovelace & Babbage and the Creation of the 1843 'Notes'." *IEEE Annals of the History of Computing* (IEEE) 16 - 26. doi:10.1109/MAHC.2003.1253887.

Gal, Peter, Giuseppe Nicoletti, Theodore Renault, Stéphane Sorbe, and Christina Timiliotis. 2019. "Digitalisation and productivity: In search of the holy grail – Firm-level empirical evidence from EU countries." *OECD Economics Department Working Papers* (OECD Publishing) (1533): 1-67. Accessed December 28, 2022. doi:10.1787/5080f4b6-en.

Gandy, Robin. 2002. *Human versus mechanical intelligence.* Vol. 1, in *Machines and thought: The Legacy of Alan Turing*, edited by Peter Millican and Andy Clark, 125 - 136. Oxford: Oxford University Press.

Garnelo, Marta, and Murray Shanahan. 2019. "Reconciling deep learning with symbolic artificial intelligence: representing objects and relations." *Current Opinion in Behavioral Sciences* 29: 17-23. doi:10.1016/j.cobeha.2018.12.010.

Gaut, Berys. 2010. "The Philosophy of Creativity." *Philosophy Compass* 5 (12): 1034 - 1046. Accessed December 27, 2022. doi:10.1111/j.1747-9991.2010.00351.

Germanidis, Anastasis. 2020. "Building a web-based real-time video editing tool with machine learning." *Runway Blog.* November 17. Accessed April 20, 2023. https://runwayml.com/blog/building-a-web-based-real-time-video-editing-tool-with-machine-learning/.

Ghuloum, Mohamed Haider. 2015. "Netanyahu's Good Arab Proverb: The dogs bark but will the caravan move on?" *Arabia Deserta: MENA Focus.* March 1. Accessed December 5, 2022. http://arabiadeserta.com/2015/03/netanyahus-good-arab-proverb-the-dogs-bark-but-will-the-caravan-move-on/.

Gibbs-Bravo, Andrew. 2022. "Paraphrase Generation Using Deep Reinforcement Learning – Thought Leaders." *Unite.ai.* December 9. Accessed Janaury 6, 2023. https://www.unite.ai/paraphrase-generation-using-deep-reinforcement-learning-thought-leaders/.

Gibbs-Bravo, Andrew. 2019. "ParaPhrasee: Paraphrase Generation using Deep Reinforcement Learning." Project Report | Thesis, City University London, London, 173. https://github.com/gibbsbravo/ParaPhrasee/blob/master/Paraphrase%20Generation%20using%20Deep%20Reinforcement%20Learning.pdf.

gling. 2023. *Creator, get your time back.* Accessed April 12, 2023. https://www.gling.ai/.

Gonçalves, Bernardo. 2022. "Can machines think? The controversy that led to the Turing test." *AI and Society* 1-11. doi:10.1007/s00146-021-01318-6.

Goodfellow, Ian J., Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. *Generative Adversarial Networks.* Research findings, arXiv:1406.2661, 1-9. doi:10.48550/arXiv.1406.2661.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning.* Cambridge, MA: MIT Press. Accessed October 2022.

Google. 2023. "Search." January 18. https://www.google.com/search?q=AI+News.

—. 2019. *YouTube-8M Segments Dataset.* September 4. Accessed March 24, 2023. https://research.google.com/youtube8m/.

Gottschall, Jonathan. 2012. *The Storytelling Animal: How Stories Make Us Human.* New York: Houghton Mifflin Harcourt Publishing Company.

Griffin, Emma. 2013. *Liberty's Dawn: A People's History of the Industrial Revolution.* New Haven, Conneticut: Yale University Press. Accessed November 2022.

Hachman, Mark. 2022. "Hands on with Microsoft Designer, an AI art masterpiece." *PCWorld.* IDG Communications, Inc. November 30. Accessed December 8, 2022. https://www.pcworld.com/article/1419066/microsoft-designer-is-almost-a-masterpiece-already.html.

Hajj, Mustapha El. 2023. "The Fascinating Evolution of Chat GPT." *Medium.* January 3. Accessed January 4, 2023. https://medium.com/life-3-0-journey/the-fascinating-evolution-of-chat-gpt-f4218b36001a.

Hale, James. 2019. "More Than 500 Hours Of Content Are Now Being Uploaded To YouTube Every Minute." *Tubefilter.* May 7. https://www.tubefilter.com/2019/05/07/number-hours-video-uploaded-to-youtube-per-minute/.

Harrison, Maggie. 2022. "There's a Problem With That AI Portrait App: It Can Undress People Without Their Consent." *Futurism.* December 7. Accessed February 9, 2023. https://futurism.com/ai-portrait-app-nudes-without-consent.

Hartree, Douglas R. 1950. *Calculating Instruments and Machines.* Cambridge: Cambridge University Press.

Hartree, Douglas Rayner. 1946. "The 'Electronic Brain'." *The Times*, November 7: 5.

Hatch, Mark. 2013. *The Maker Movement Manifesto: Rules for Innovation in the New World of Crafters, Hackers, and Tinkerers.* New York, NY: McGraw Hill. Accessed November 2022.

Hautamäki, Antti, Juha Leppänen, Roope Mokka, and Aleksi Neuvonen. 2017. *From pause to play. Work and income in the next era.* Memorandum, Sitra Fund, Helsinki: Sitra, 1-38. https://www.sitra.fi/app/uploads/2017/01/From_pause_to_play_Sitra_Demos.pdf.

Hayles, N. Katherine. 2016. "Electronic Literature: What Is It?" Chap. 12a in *Doing Digital Humanities: Practice, Training, Research*, edited by Constance Crompton, Richard J Lane and Ray Siemens, 197-226. Routledge.

Heaven, Will Douglas. 2023. "The original startup behind Stable Diffusion has launched a generative AI for video." *MIT Technology Review.* February 6. Accessed February 8, 2023. https://www.technologyreview.com/2023/02/06/1067897/runway-stable-diffusion-gen-1-generative-ai-for-video/.

Helyer, Ruby. 2022. "What Is Adobe Sensei? How This Artificial Intelligence Tool Helps Creators." *Make Use Of.* June 11. Accessed April 20, 2023. https://www.makeuseof.com/what-is-adobe-sensei/.

Hern, Alex. 2019. "New AI fake text generator may be too dangerous to release, say creators." *The Guardian*, February 14. Accessed December 5, 2022. https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction.

Hernandez, Sergio Poo, Vadim Bulitko, and Emilie St. Hilaire. 2014. "Emotion-Based Interactive Storytelling with Artificial Intelligence." *Proceedings of the Tenth Annual AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE).* Raleigh, North Carolina: AAAI Publications. 1-7. https://www.aaai.org/ocs/index.php/AIIDE/AIIDE14/paper/view/8982/8941.

Hertzmann, Aaron. 2018. "Can Computers Create Art?" Edited by Dr. Frederic Fol Leymarie and G. W. Smith. *Arts: Special Issue The Machine as Artist (for the 21st Century)* 7 (2): 1-25. Accessed December 2022. doi:10.3390/arts7020018.

Hewitt, Carl. 1969. "Planner: A Language for Proving Theorems in Robots." Edited by Donald Walker and L. Norton. *Proceedings of IJCAI-69* (International Joint Conference on Artificial Intelligence) 295-301.

Hipkins, Rosemary, and Bronwen Cowie. 2016. "The Sigmoid Curve as a Metaphor for Growth and Change." *Teachers and Curriculum* 16 (2): 3-9.

Ho, Jonathan, Ajay Jain, and Pieter Abbeel. 2020. "Denoising Diffusion Probabilistic Models." *Advances in Neural Information Processing Systems* 33: 6840-6851. doi:10.48550/arXiv.2006.11239.

Hollister, Sean. 2023. "'AI is going to reinvent how you do everything on Windows,' says Microsoft's Windows boss." *The Verge.* Janaury 5. Accessed January 16, 2023. https://www.theverge.com/2023/1/4/23539873/microsoft-panos-panay-ai-windows-amd-ces-2023.

Hong, Joo-Wha, and Nathaniel Ming Curran. 2019. "Artificial Intelligence, Artists, and Art: Attitudes Toward Artwork Produced by Humans vs. Artificial Intelligence." *ACM Transactions on Multimedia Computing, Communications, and Applications* 15 (2s): 1-16. Accessed December 20, 2022. doi:10.1145/3326337.

Howard, T. 1997. "Intelligence test." *Personal Computer World*, August: 234-5.

Huang, Sonya, and Pat Grady. 2022. "Generative AI: A Creative New World." *Sequoia Capital.* Accessed December 2, 2022. https://www.sequoiacap.com/article/generative-ai-a-creative-new-world/.

Hutson, Matthew. 2022. "TAUGHT TO THE TEST AI software clears high hurdles on IQ tests but still makes dumb mistakes. Can better benchmarks help?" *Science* 376 (6593): 570 - 573. doi:10.1126/science.abq7853.

Iku, Data. 2021. "Dartmouth Summer Research Project: The Birth of Artificial Intelligence." *The History of Data Science.* September 30. Accessed February 20, 2023. https://www.historyofdatascience.com/dartmouth-summer-research-project-the-birth-of-artificial-intelligence/.

ImageNet. 2021. *ImageNet.* March 11. Accessed March 14, 2023. https://www.image-net.org/index.php.

Inge, William Ralph. 1929. *Assessments and Anticipation.* London: Cassell and Company. Accessed 2022.

Intel. 2022. *Intel® Deep Learning Streamer.* Intel Corporation. Accessed April 24, 2023. https://dlstreamer.github.io/.

International Data Corporation. 2022. "Generative AI: What does it mean in the Enterprise?" *IDC Blog.* December 21. Accessed February 14, 2023. https://blogs.idc.com/2022/12/21/generative-ai-what-does-it-mean-in-the-enterprise/.

Jackson, Emily. 1900. *A history of hand-made lace : Dealing with the origin of lace, the growth of the great lace centres, the mode of manufacture, the methods of distinguishing and the care of various kinds of lace.* London: L. Upcott Gill. Accessed December 2022. https://ia802803.us.archive.org/4/items/historyofhandmad00jack/historyofhandmad00jack_bw.pdf.

Jackson, Tony E. 2017. "Imitative Identity, Imitative Art, and AI: Artificial Intelligence." *Mosaic: An Interdisciplinary Critical Journal* (University of Manitoba Stable) 50 (2): 47-63. Accessed December 2022. https://www.jstor.org/stable/45158927.

Javed, Ali, Aun Irtaza, Hafiz Malik, Muhammad Mahmood, and Syed Adnan. 2019. "A Multimodal Framework based on Audio-Visual Features for Summarization of Cricket Videos." *IET Image Processing* (IETDL) (13): 1-8. Accessed March 2022. doi:10.1049/iet-ipr.2018.5589.

Javed, Ali, Khalid Malik, Aun Irtaza, and Hafiz Malik. 2020. "A decision tree framework for shot classification of field sports videos." *The Journal of Supercomputing* (Springer Science+Business Media, LLC) 76 (5). Accessed March 2022. doi:10.1007/s11227-020-03155-8.

Jefferson, Geoffrey. 1949. "The Mind of Mechanical Man." *British Medical Journal* 1 (4616): 1105–1110. doi:10.1136/bmj.1.4616.1105.

Jenkins, Henry. 2007. "Introduction: “Worship at the Altar of Convergence” - A New Paradigm for Understanding Media Change." *dossier* La culture de la convergence : 31-36. Accessed September 23, 2021. http://documents.irevues.inist.fr/bitstream/handle/2042/23560/2007_21_31.pdf.

Ji, Anna, interview by Drew Keller. 2022. *ClipChamp Co-Founder* (February 9).

Jordan, Jeremy. 2017. "Neural networks: representation." *Jeremy Jordan.* June 28. Accessed March 6, 2023. https://www.jeremyjordan.me/intro-to-neural-networks/.

Kafai, Yasmin, and Jane Margolis. 2016. "Changing the Face of Computing—One Stitch at a Time." *The MIT Press.* October 11. Accessed February 2, 2023. https://mitpress.mit.edu/celebrating-ada-lovelace/.

Karnofsky, Holden. 2021. "Forecasting Transformative AI, Part 1: What Kind of AI." *Cold Takes.* August 10. Accessed December 14, 2022. https://www.cold-takes.com/transformative-ai-timelines-part-1-of-4-what-kind-of-ai/.

Kaur, Navleen, Supriya Lamba Sahdev, Monika Sharma, and Laraibe Siddiqui. 2020. "Banking 4.0: ‘The Influence of Artificial Intelligence on the Banking Industry & How AI Is Changing the Face of Modern Day Banks’." *International Journal of Management* 11 (6): 577-585. Accessed December 2022. doi:10.34218/IJM.11.6.2020.049.

Kautz, Henry. 2020. "The Third AI Summer, Henry Kautz, AAAI 2020 Robert S. Engelmore Memorial Award Lecture." *34th Annual Meeting of the Association for the Advancement of Artificial Intelligence.* New Yotu: YouTube, February 10. https://youtu.be/_cQITY0SPiw.

Keen, Martin. 2023. "Why Large Language Models Hallucinate." *IBM Technology.* YouTube, April 20. https://youtu.be/cfqtFvWOfg0.

Keynes, John Maynard. 1931. *Essays in persuasion.* London: MacMillan and Co. Accessed December 2022.

King, Mike. 2002. "Artificial consciousness - artificial art." Chap. 13 in *Computers and Art*, edited by Stuart Mealing, 143-158. Portland, OR: Intelect. Accessed December 2022.

Kleon, Austin. 2012. *Steal Like an Artist: 10 Things Nobody Told You About Being Creative.* New York, New York: Workman Publishing. Accessed 2013.

Knuth, Donald E. 1996. "Forward." In *A=B*, by Marko Petkovsek, Herbert S. Wilf and Doron Zeilberger, xi-xii. Boca Raton, FL: CRC Press / Taylor & Francis Group. Accessed December 2022. https://www2.math.upenn.edu/~wilf/AeqB.pdf.

Kolekar, Maheshkumar H., and Somnath Sengupta. 2006. "A Hierarchical Framework for Generic Sports Video Classification." *Computer Vision–ACCV 2006: 7th Asian Conference on Computer Vision* (Springer Berlin Heidelberg) II (7): 633-642.

Krogh, Anders. 2008. "What are artificial neural networks?" *Nature biotechnology* 26 (2): 195-197. https://www.nature.com/articles/nbt1386.

Krugman, Paul. 2023. "Making America the Opposite of Great." *New York Times.* January 5. Accessed January 5, 2023. https://www.nytimes.com/2023/01/05/opinion/america-kevin-mccarthy-great.html.

Kurt, Deniz E. 2018. "Artistic Creativity in Artificial Intelligence." Nijmegen: Radboud University , April. 1-87. Accessed December 19, 2022. https://theses.ubn.ru.nl/handle/123456789/5631.

Kurzweil, Ray. 1999. *The age of spiritual machines: When computers exceed human intelligence.* Penguin.

—. 2001. "The Law of Accelerating Returns." *Kurzweil.* March 7. Accessed April 27, 2023. https://www.kurzweilai.net/the-law-of-accelerating-returns.

Kurzweil, Ray. 2014. "The Singularity is Near - Excerpt." In *Ethics and Emerging Technologies*, edited by Ronald L. Sandler, 393–406. London: Palgrave Macmillan. Accessed December 5, 20223. doi:10.1057/9781137349088_26.

—. 2005. *The Singularity is Near.* Penguin Press.

Kurzweil, Ray. 1999b. "Spiritual machine." *Research & Development; Highlands Ranch*, June: 14-18.

Kushner, David. 2009. "When Man & Machines Merge." *Rolling Stone*, February 19: 57 - 61. https://issuu.com/i48998/docs/kurzweil_in_rolling_stone.

LAION. 2023. *LAION Home Page.* March. Accessed March 13, 2023. https://laion.ai/.

Lardinois, Frederic. 2022. "IBM unveils its 433 qubit Osprey quantum computer." *TechCrunch.* November 9. Accessed January 4, 2023. https://techcrunch.com/2022/11/09/ibm-unveils-its-433-qubit-osprey-quantum-computer/.

LeCun, Yann. 2017. "Yann LeCun - Power & Limits of Deep Learning." *AI and the Future of Work.* Cambridge, Massachusetts: YouTube, November 19.

Leong, Chee Wee, and Rada Mihalcea. 2011. "Measuring the semantic relatedness between words and images." *Proceedings of the Ninth International Conference on Computational Semantics.* Oxford, UK: IWCS. 185-194.

Lin, David Chuan-En, Anastasis Germanidis, Cristóbal Valenzuela, Yining Shi, and Nikolas Martelaro. 2021. *Soundify: Matching Sound Effects to Video.* Research findings, arXiv:2112.09726, 1-7. doi:10.48550/arXiv.2112.09726.

Lovelace, Ada Augusta, Countess of. 1842. "Sketch of The Analytical Engine Invented by Charles Babbage - Notes." *Bibliothèque Universelle de Genève* 82. https://www.fourmilab.ch/babbage/sketch.html.

Lovett, William. 1876. *Life and Struggles of William Lovett, in His Pursuit of Bread, Knowledge, and Freedom.* Knopf: London.

Luntz, Stephen. 2022. "New Supercomputer First To Officially Break The Quintillion Calculations A Second Barrier." *IFL Science.* June 1. Accessed March 16, 2023. https://www.iflscience.com/new-supercomputer-first-to-officially-break-the-quintillion-calculations-a-second-barrier-63902.

Maeda, John. 2022. "Design and artificial intelligence: hype or hope?" *Microsoft Design Week 2022.* Redmond, Washington: Microsoft, December 5. Accessed December 6, 2022.

Market.us. 2023. *AI Video Generator Market Size, Share, Trends, Forecast - 2032.* Market forecast, Market.us, 362. https://market.us/report/ai-video-generator-market/#overview.

Marr, Bernard. 2021. "What is Weak (Narrow) AI? Here Are 8 Practical Examples." *Pulse.* May 27. Accessed March 1, 2023. https://www.linkedin.com/pulse/what-weak-narrow-ai-here-8-practical-examples-bernard-marr/.

Martindale, Colin. 1999. "Biological Bases of Creativity." In *Handbook of creativity 2* , 137-152. Cambridge University press.

Massey, Douglas S. 2002. "A brief history of human society: The origin and role of emotion in social life." *American Sociological Review* 67 (1): 1-29. Accessed 2017. http://www.wimvanmourik.eu/logica_gevoel/theorie/a_brief_history_of_human_society.pdf.

Massumi, Brian. 1992. "'Anywhere you want to be': An introduction to fear." In *Deleuze and the Transcendental Unconscious*, by Joan (ed.) Broadhurst, edited by Joan (ed.) Broadhurst, 175-215. Warwick: Warwick Journal of Philosophy.

McCarthy, John. 2007. "From here to human-level AI." *Artificial Intelligence* 171 (18): 1174-1182. Accessed 2022. doi:10.1016/j.artint.2007.10.009.

McCarthy, John. 2007. *What is Artificial Intelligence?* Article, Computer Science Department, Stanford University, Stanford: Stanford University, 15. Accessed December 2022. http://jmc.stanford.edu/articles/whatisai/whatisai.pdf.

McCarthy, John, Marvin L Minsky, Nathaniel Rochester, and Claude E Shannon. 2006. "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence: August 31, 1955." *AI Magazine* 27 (4): 12-14.

McCorduck, Pamela. 2004. *Machines who think: A personal inquiry into the history and prospects of artificial intelligence.* Natick, MA: Taylor & Francis Group.

McCulloch, Warren S., and Walter Pitts. 1943. "A Logical Calculus of the Ideas Immanent in Nervous Activity." *The bulletin of mathematical biophysics* 5: 115–133.

McDermott, Drew. 2014. "What was Alan Turing's imitation game?." *The critique, as part of an issue on Turing's imitation game.* December 31. Accessed February 24, 2023. http://www.thecritique.com/articles/what-was-alan-turings-imitation-game/.

McGinnis, Devon. 2020. "Meet the Three Industrial Revolutions." *The 360 Blog.* October 27. Accessed January 30, 2023. https://trailhead.salesforce.com/content/learn/modules/learn-about-the-fourth-industrial-revolution/meet-the-three-industrial-revolutions.

McKinsey. 2023. "What is generative AI?" *McKinsey & Company.* January 19. Accessed February 14, 2023. https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai.

Meng, Jianliang, Junwei Zhang, and Haoquan Zhao. 2012. "Overview of the Speech Recognition Technology." *Fourth International Conference on Computational and Information Sciences.* Chongqing, China: Institute of Electrical and Electronics Engineers. 199-202. doi:10.1109/ICCIS.2012.202.

Mercado, Gustavo. 2022. *The Filmmaker's Eye.* 2nd. Focal Press. doi:10.4324/9781315770857.

Merchant, Brian. 2019. "There's an Automation Crisis Underway Right Now, It's Just Mostly Invisible." *Gizmodo.* October 11. Accessed Janaury 24, 2023. https://gizmodo.com/the-trickle-down-disaster-of-automation-1838974516.

Merzmensch. 2022. "Artstation: "AI art free zone"…." *Medium.* December 14. Accessed February 9, 2023. https://medium.com/merzazine/artstation-ai-art-free-zone-b042576df592.

—. 2022. "Prompt Design for DALL·E 2: Series." *Medium.* June 5. Accessed Janaury 25, 2023. https://medium.com/merzazine/prompt-design-for-dall-e-2-series-5e50cec59d08.

Meta. 2022. *Make-A-Video.* Accessed March 24, 2023. https://makeavideo.studio/.

Metz, Cade. 2023. "Instant Videos Could Represent the Next Leap in A.I. Technology." *The New York Times*, April 4: 1-4. https://www.nytimes.com/2023/04/04/technology/runway-ai-videos.html.

Metz, Cade, and Karen Weise. 2023. "Microsoft Bets Big on the Creator of ChatGPT in Race to Dominate A.I." *The New York Times.* Janaury 12. Accessed January 13, 2023. https://www.nytimes.com/2023/01/12/technology/microsoft-openai-chatgpt.html.

Microsoft. 2023. "Custom Vision." *Azure.* Accessed April 24, 2023. https://azure.microsoft.com/en-us/products/cognitive-services/custom-vision-service/.

—. 2023. *Responsible AI.* Accessed February 15, 2023. https://www.microsoft.com/ai/responsible-ai.

Midha, Vinay Kumar, and Dakuri Arjun. 2017. "Spun bonding Technology and Fabric Properties: a Review." *Journal of Textile Engineering & Fashion Technology* 1 (4): 126-133. doi:10.15406/jteft.2017.01.00023.

Midhu, K., and N. K. Anantha Padmanabhan. 2018. "Highlight Generation of Cricket Match Using Deep Learning." *Computational Vision and Bio Inspired Computing* (Springer International Publishing) 925-936. doi:10.1007/978-3-319-71767-8_79.

Mindell, David A. 1995. "Anti-Aircraft Fire Control and the Development of Integrated Systems at Sperry, 1925–1940." *IEEE Control Systems Magazine* 15 (2): 108–13. doi:10.1109/37.375318.

Modis, Theodore. 2006. "Book Review and Discussion - The Singularity is Near." *Technological Forecasting and Social Change* 73 (2): 95-127. doi:10.1016/j.techfore.2005.12.003.

Mohsin, Maryam. 2021. "10 YouTube Statistics Every Marketer Should Know." *Oberlo.* January 25. https://www.oberlo.com/blog/youtube-statistics.

Molad, Eyal, Eliahu Horwitz, Dani Valevski, Alex Rav Acha, Yossi Matias, Yael Pritch, Yaniv Leviathan, and Yedid Hoshen. 2023. *Dreamix: Video Diffusion Models are General Video Editors.* Research findings, arXiv:2302.01329, 1-18. doi:10.48550/arXiv.2302.01329.

Mollick, Ethan. 2022. "ChatGPT Is a Tipping Point for AI." *Harvard Business Review.* December 14. Accessed Janaury 13, 2023. https://hbr.org/2022/12/chatgpt-is-a-tipping-point-for-ai.

Moore, Gordon E. 1965. "Cramming more components onto integrated circuits." *Electronics* 38 (8): 1-4.

Moruzzi, Caterina. 2020. "Learning through creativity: How creativity can help machine learning achieving deeper understanding." *Rivista Italiana Di Filosofia Del Linguaggio* 14 (2): 1-12. Accessed December 26, 2022. doi:10.4396/AISB201904.

Mou, Yi, and Kun Xu. 2017. "The media inequality: Comparing the initial human-human and human-AI social interactions." *Computers in Human Behavior* 72: 432–440. Accessed December 2022. doi:10.1016/j.chb.2017.02.067.

Mui, Chunka. 2012. "How Kodak Failed." *Forbes.* January 18. Accessed April 12, 2023. https://www.forbes.com/sites/chunkamui/2012/01/18/how-kodak-failed/.

Nass, Clifford, and Youngme Moon. 2000. "Machines and mindlessness: Social responses to computers." *Journal of Social Issues* 56 (1): 81–103. Accessed December 2022. doi:10.1111/0022-4537.00153.

Newman, Maxwell Herman Alexander. 1955. "Alan Mathison Turing, 1912–1954." *Biographical Memoirs of Fellows of the Royal Society* 253–263. https://www.jstor.org/stable/769256.

Norman, Donald A., and David E. Rumelhart. 1981. "The LNR approach to human information processing." *Cognition* 10 (1): 235–240. doi:10.1016/0010-0277(81)90051-2.

Ong, Thuy. 2017. "Amazon's New Algorithm Designs Clothing by Analyzing a Bunch of Pictures." *The Verge.* August 24. Accessed December 12, 2022. https://www.theverge.com/2017/8/24/16195858/amazon-ai-fashion-designer.

Open AI. 2022. *ChatGPT.* December 15. Accessed January 4, 2023. https://chat.openai.com/chat.

O'Riordan, Lizzy. 2022. "We Take a Trip to Cluny Lace, the Last Traditional Lace Maker in the UK." *Leftlion.* March 19. Accessed December 19, 2022. https://www.leftlion.co.uk/read/2022/march/cluny-lace-ilkeston-the-last-lace-maker/.

Ouyang, Long, Xu Jiang Jeff Wu, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, et al. 2022. *Training language models to follow instructions with human feedback.* Research findings, OpenAI, arXiv, 1-68. Accessed December 8, 2022. doi:10.48550/arXiv.2203.02155.

Owens, Mackubin Thomas. 2001. "Marines Turned Soldiers....The Corps vs. the Army." *National Review Online.* December 10. Accessed December 8, 2022. http://www.nationalreview.com/comment/comment-owens121001.shtml.

Paetzhold, Megan. 2022. "Will Dalle-E the AI artist take my job?" *New York Magazine.* September 4. Accessed December 19, 2022. https://nymag.com/intelligencer/article/will-dall-e-ai-artist-take-my-job.html.

Palliser, Bury. 1869. *A History of Lace.* Second. London: Sampson, Low, Son and Marston. Accessed December 2022.

Pankin, Jeff. 2013. "Schema Theory." Cambridge, MA, Fall. http://web.mit.edu/pankin/www/Schema_Theory_and_Concept_Formation.pdf.

Parsons, Talcott. 1970. "Theory in the Humanities and Sociology." *Daedalus* (The MIT Press) 99 (2): 495-523. Accessed August 13, 2021. https://www.jstor.org/stable/20023955.

Pedersen, Morten Axel, Kristoffer Albris, and Nick Seaver. 2021. "The Political Economy of Attention." *Annual Review of Anthropology* 50: 309-325. Accessed November 1, 2021. https://doi.org/10.1146/annurev-anthro-101819-110356.

Peebles, William, Jun-Yan Zhu, Richard Zhang, Antonio Torralba, Alexei A. Efros, and Eli Shechtman. 2022. "GAN-Supervised Dense Visual Alignment." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* CVPR. 13470-13481. doi:10.48550/arXiv.2112.05143.

PEGA. 2022. *The future of work: New perspectives on disruption & transformation.* Technical whitepaper, Pegasystems, 23. https://www.pega.com/the-future-of-work.

Pei, Soo-Chang, and Fan Chen. 2003. "Semantic scenes detection and classification in sports videos." *Proceedings of IPPR Conference on Computer Vision, Graphics and Image Processing (CVGIP).* Kinmen, ROC : CVGIP. 210-217.

Peiravi, Mehdi. 2015. *The Design and Implementation of an Adaptive Chess Game.* Research findings, San Bernardino: California State University, 106.

Pfau, David, and Oriol Vinyals. 2016. *Connecting Generative Adversarial Networks and Actor-Critic Methods.* Research findings, arXiv:1610.01945. doi:10.48550/arXiv.1610.01945.

Philpott, Eric. 2023. "NAB 2023: Introducing Text-Based Editing in Premiere Pro, Properties panel in After Effects, and much more." *Adobe Blog.* April 13. Accessed April 20, 2023. https://blog.adobe.com/en/publish/2023/04/13/nab-2023-introducing-text-based-editing-premiere-pro-properties-panel-after-effects-more.

Pillai, Gokul. 2023. "Slumdog Millionaires - Series of 7." Midjourney. *Instagram.* https://www.instagram.com/p/CqvxGHwyyf1/.

Popova, Maria. 2011. "Networked Knowledge and Combinatorial Creativity." *The Marginalian.* May. Accessed December 12, 2022. https://www.themarginalian.org/2011/08/01/networked-knowledge-combinatorial-creativity/.

2022. *How AI Image Generators Work (Stable Diffusion / DALL·E).* Directed by Sean Riley. Produced by Computerphile. Performed by Michael Pound. YouTube. Accessed November 7, 2022. doi:https://youtu.be/1CIpzeNxIhU.

Proudfoot, Diane. 2015. "Mocking AI Panic." *IEEE Spectrum.* June 30. Accessed December 5, 2022. https://spectrum.ieee.org/mocking-ai-panic.

Proudfoot, Diane. 2020. "Rethinking Turing's Test and the Philosophical Implications." *Minds and Machines* 30 (4): 487–51. doi:10.1007/s11023-020-09534-7.

Pueringer, Niko. 2022. "Is This The Death of VFX?" *Corridor Crew.* Los Angeles, California: YouTube, September 25. Accessed December 13, 2022. https://youtu.be/W4Mcuh38wyM.

Pulliam-Moore, Charles. 2023. "The Writers Guild of America likens AI-generated content to plagiarism." *The Verge.* March 22. Accessed April 7, 2023. https://www.theverge.com/2023/3/22/23651804/wga-union-chatgpt-ai-tools-proposal.

Puschak, Evan. 2022. "The Real Danger Of ChatGPT." *Nerdwriter1.* YouTube, December 30. https://youtu.be/AAwbvGywdOc.

qFiasco, Flash. 2018. "Deep Thinking, Where Machine Intelligence Ends and Human Creativity Begins, Garry Kasparov, Mig Greengard, John Murray." *Artificial Intelligence* (Elsevier) 260: 36–41. Accessed December 2022. doi:10.1016/j.artint.2018.04.001.

Quyang, Long, interview by David Burch. 2023. *OpenAI on Reinforcement Learning With Human Feedback* Arize, (Janaury 19). https://arize.com/blog/openai-on-rlhf/.

Rabiner, L., and B. Juang. 1986. "An introduction to hidden Markov models." *IEEE ASSP Magazine*, January: 4-16. doi:10.1109/MASSP.1986.1165342.

Radford, Alec, Ilya Sutskever, Jong Wook Kim, Gretchen Krueger, and Sandhini Agarwal. 2021. *CLIP: Connecting text and images.* Research findings, OpenAI. https://openai.com/research/clip.

Radford, Alec, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2022. *Robust Speech Recognition via Large-Scale Weak Supervision.* Research results, Open AI Labs, arXiv:2212.04356, 1-28. Accessed December 5, 2022. doi:10.48550/arXiv.2212.04356.

Ramesh, Aditya, interview by Cleo Abram. 2022. *The REAL fight over AI art* YouTube. New York. September 14. Accessed September 19, 2022. https://www.youtube.com/watch?v=NiJeB2NJy1A.

Ramesh, Aditya, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. *Zero-Shot Text-to-Image Generation.* arXiv preprint, arXiv:2102.12092, 1-20. doi:10.48550/arXiv.2102.12092.

Ramesh, Aditya, Mikhail Pavlov, Gabriel Goh, Scott Gray, Mark Chen, Rewon Child, Vedant Misra, et al. 2021b. *DALL·E: Creating images from text.* Research findings, Research, OpenAI, OpenAI, 1-20. https://openai.com/research/dall-e.

Ramesh, Aditya, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. *Hierarchical Text-Conditional Image Generation with CLIP Latents.* Research findings, Ithaca, NY: arXiv. Accessed November 21, 2022. doi:10.48550/ARXIV.2204.06125.

—. 2022b. *How DALL·E 2 Works.* Accessed December 6, 2022. http://adityaramesh.com/posts/dalle2/dalle2.html.

Raventós, A, R Quijada, Luis Torres, and Francesc Tarrés. 2015. "Automatic summarization of soccer highlights using audio-visual descriptors." *SpringerPlus* 4 (301): 1–19. Accessed March 2022. doi:10.1186/s40064-015-1065-9.

RCA. 1958. "Studio: The World--NBC Introduces 'Chroma-Key' to Extend Scope of TV Settings." *Electronic Age*, January: 8-9, 30. https://worldradiohistory.com/Archive-Radio-Age/Electronic-Age-1958-Winter.pdf.

Read, Paul, and Mark-Paul Meyer. 2000. *Restoration of Motion Picture Film.* Butterworth-Heinemann.

Reagan, Andrew J, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds. 2016. "The emotional arcs of stories are dominated by six basic shapes." *The European Physical Journal Data Science* 5: 1-12. doi:10.1140/epjds/s13688-016-0093-1.

2017. *Deeply Artificial Trees.* Directed by Alex Reben. https://areben.com/project/deeply-artificial-trees/.

Reese, Byron. 2018. *The Fourth Age: Smart Robots, Conscious Computers, and the Future of Humanity.* New York: Atria Books. Accessed December 2022.

Reese, Hope. 2020. "A Brief History of the Rubik's Cube." *Smithsonian Magazine*, September 25. https://www.smithsonianmag.com/innovation/brief-history-rubiks-cube-180975911/.

Rifkin, Jeremy. 1995. *The End of Work.* New York, New York: G.P. Putnam's Sons. Accessed December 2022. http://pinguet.free.fr/rifkin1995.pdf.

Roach, John. 2022. "From Hot Wheels to handling content: How brands are using Microsoft AI to be more productive and imaginative." *The AI Blog.* Microsoft. October 12. Accessed December 5, 2022. https://blogs.microsoft.com/ai/from-hot-wheels-to-handling-content-how-brands-are-using-microsoft-ai-to-be-more-productive-and-imaginative/.

Rodgers, Ornaith, and Labhaoise Ni Dhonnchadha. 2018. "Digital video creation in the LSP classroom." Edited by Ana Gimeno. *The EUROCALL Review* (European Association for Computer-Assisted Language Learning) 26 (1): 43-58. doi:https://doi.org/10.4995/eurocall.2018.9666.

Rogers, Everett. 2003. *Diffusion of Innovations.* Fifth. The Free Press. Accessed December 2022. http://www.lamolina.edu.pe/postgrado/pmdas/cursos/innovacion/lecturas/Obligatoria/17%20-%20Rogers%201995%20cap%206.pdf.

Rogers, Henry W. 1883. "Sarony v. Burrow-Giles Lithographic Co." *University of Michigan Law School.* October. Accessed April 21, 2023. https://repository.law.umich.edu/articles/2142/.

Roose, Kevin. 2022. *Futureproof: 9 Rules for Humans in the Age of Automation.* 2. New York, New York: Random House Trade Paperbacks. Accessed December 2022.

—. 2022. "The Brilliance and Weirdness of ChatGPT." *The New York Times.* December 5. Accessed January 4, 2023. https://www.nytimes.com/2022/12/05/technology/chatgpt-ai-twitter.html.

—. 2022. "A Coming-Out Party for Generative A.I., Silicon Valley's New Craze." *The New York Times*, October 21. Accessed December 2, 2022. https://www.nytimes.com/2022/10/21/technology/generative-ai.html.

—. 2022. "A.I.-Generated Art Is Already Transforming Creative Work." *The New York Times*, October 21. Accessed December 2, 2022. https://www.nytimes.com/2022/10/21/technology/ai-generated-art-jobs-dall-e-2.html.

Rose, Sonya O. 1992. *Limited Livelihoods: Gender and Class in Nineteenth-Century England.* London: Routledge.

Rosenbaum, Ron. 2013. "What Turned Jaron Lanier Against the Web?" *Smithsonian Magazine.* January. Accessed January 2, 2023. https://www.smithsonianmag.com/innovation/what-turned-jaron-lanier-against-the-web-165260940/.

Rosenblatt, Frank. 1961. *Principles of neurodynamics. Perceptrons and the theory of brain mechanisms.* Research findings, Ithaca, NY: Cornell Aeronautical Lab Inc.

Rosenblatt, Frank. 1957. "The perceptron: A probabilistic model for information storage and organization in the brain." *Psychological Review* 65 (6): 386–408. doi:10.1037/h0042519.

Ruby, Daniel. 2022. "ChatGPT Statistics for 2023: Comprehensive Facts and Data." *Demand Sage.* December 29. Accessed January 17, 2023. https://www.demandsage.com/chatgpt-statistics/.

Runway Research. 2023. *Gen-1: The Next Step Forward for Generative AI.* February 5. Accessed February 8, 2023. https://research.runwayml.com/gen1.

Runway. 2022. "Text to video." *Twitter.* RunwayML, September 9. https://twitter.com/runwayml/status/1568220303808991232?s=20&t=vm7fUFRnOx0qBMrD4dsr_w.

Russell, Stuart J., and Peter Norvig. 2003. *Artificial Intelligence: A Modern Approach.* Second. Upper Saddle River, NJ: Prentice Hall.

Samuel, Sigal. 2022. "A new AI draws delightful and not-so-delightful images." *Vox.* April 14. Accessed November 21, 2022. https://www.vox.com/future-perfect/23023538/ai-dalle-2-openai-bias-gpt-3-incentives.

Sandler, Ronald L. 2014. "Part VI: Robotics and Artificial Intelligence." In *Ethics and Emerging Technologies*, 347-348. London: Palgrave Macmillan. Accessed December 5, 2022. doi:10.1057/9781137349088_26.

Sawyer, R. Keith. 2012. *Explaining Creativity: The Science of Human Innovation.* New York, NY: Oxford University Press. Accessed November 2022.

Scherer, F. M. 1965. "Invention and Innovation in the Watt-Boulton Steam-Engine Venture." *Technology and Culture* 6 (2): 165-187. https://www.jstor.org/stable/3101072.

Schroer, Alyssa, and Ellen Glover. 2022. "What Is Artificial Intelligence? How Does AI Work?." *Built In.* September 19. Accessed December 20, 2022. https://builtin.com/artificial-intelligence.

Schubert, Christian. 2013. "How to evaluate creative destruction: reconstructing Schumpeter's approach." *Cambridge Journal of Economics* 37 (2): 227–250. doi:10.1093/cje/bes055.

Schuhmann, Christoph, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, et al. 2022. *LAION-5B: An open large-scale dataset for training next generation image-text models.* Research Findings, arXiv:2210.08402. doi:10.48550/arXiv.2210.08402.

Schumpeter, Joseph A. 1942. *Capitalism, Socialism and Democracy.* Harper & Brothers.

Schuster, Mike, Melvin Johnson, and Nikhil Thorat. 2016. "Zero-Shot Translation with Google's Multilingual Neural Machine Translation System." *Google AI Blog.* Google Research. November 22. Accessed Janaury 26, 2023. https://ai.googleblog.com/2016/11/zero-shot-translation-with-googles.html.

Schwab, Klaus. 2017. *The Fourth Industrial Revolution.* London: Penguin Books Limited.

2022. *Best 6 Ai Tools For Designers.* Video. Directed by Ran Segall. Produced by Flux Academy. Performed by Ran Segall. YouTube. Accessed November 25, 2022. https://youtu.be/yPuP5Wbvh9s.

Sellman, Mark. 2023. "Emma Watson reads Mein Kampf on 4Chan in deepfake audio trick." *The Times of London.* Janaury 31. Accessed March 13, 2023. https://www.thetimes.co.uk/article/ai-4chan-emma-watson-mein-kampf-elevenlabs-9wghsmt9c.

Shalev-Shwartz, Shai, Shaked Shammah, and Amnon Shashua. 2017. *On a Formal Model of Safe and Scalable Self-driving Cars.* Research findings, eprint arXiv:1708.06374.

Shanahan, Murray. 2015. *The Technological Singularity.* Cambridge, MA: MIT Press.

Shannon, Claude, and John McCarthy. 1956. *Automata Studies.* Vol. 34. Princeton, NJ: Princeton University Press.

Shifman, Limor. 2014. *Memes in Digital Culture.* Cambridge, MA: The MIT Press. Accessed September 30, 2021.

Shingrakhia, Hansa, and Hetal Patel. 2021. "SGRNN-AM and HRF-DBN: a hybrid machine learning model for cricket video summarization." *Vis Comput* 1-17. doi: 10.1007/s00371-021-02111-8.

Shusterman, Neal. 2016. *Scythe (Arc of a Scythe Book 1).* Simon & Schuster.

Singer, Uriel, Adam Polyak, Thomas Hayes, Xi Yin, Jie An, Songyang Zhang, Qiyuan Hu, et al. 2022. *Make-A-Video: Text-to-Video Generation without Text-Video Data.* Research Findings, Meta, Menlo Park: arXiv, 1-13. Accessed December 2022. doi:10.48550/arXiv.2209.14792.

Smith, Thomas. 2023. "How Popular is Cricket Around the World?" *Sports Quotes and Facts.* Accessed April 21, 2023. https://sqaf.club/how-popular-is-cricket/.

Song, Yang, and Stefano Ermon. 2019. *Generative Modeling by Estimating Gradients of the Data Distribution.* Research findings, arXiv:1907.05600, 1-23. doi:10.48550/arXiv.1907.05600.

Soundraw. 2020. *Stop searching for the song you need. Create it.* February 12. Accessed Janaury 24, 2023. https://soundraw.io/.

Spacey, John. 2017. "What is Recursive Self-Improvement?" *Simplicable.* January 8. Accessed December 22, 2022. https://simplicable.com/new/recursive-self-improvement.

Srinivasan, Ragavan, interview by Hussein Mehenna. 2022. *Emerging AI Companies Are Driving A Paradigm Shift in ML* iMerit, (Janauary 25). https://imerit.net/summits/emerging-ai-companies-are-driving-a-paradigm-shift-in-ml/.

Srnicek, Nick. 2017. *Platform Capitalism.* Cambridge: Polity Press. Accessed November 2012.

Stability AI. 2022. "Stability AI Announces $101 Million in Funding for Open-Source Artificial Intelligence." *Cision PR Newswire.* October 17. Accessed December 5, 2022. https://www.prnewswire.com/news-releases/stability-ai-announces-101-million-in-funding-for-open-source-artificial-intelligence-301650932.html.

Stang, Daniel. 2020. *How AI Will Reinvent Program and Portfolio Management.* Financial analysis, Gartner Insights. https://www.gartner.com/document/3894666.

Steiner, Rupert. 2021. "Artificial Intelligence is more profound than fire, electricity, or the internet, says Google boss." *MarketWatch.* July 13. Accessed January 4, 2023. https://www.marketwatch.com/story/artificial-intelligence-is-more-profound-than-fire-electricity-or-the-internet-says-google-boss-11626202566.

Strickland, Eliza. 2022. "DALL·E 2's Failures Are the Most Interesting Thing About It OpenAI's text-to-image generator still struggles with text, science, faces, and bias." *IEEE Spectrum.* July 14. Accessed November 21, 2022. https://spectrum.ieee.org/openai-dall-e-2.

Summerlin, Will. 2022. "Productivity Gains Could Propel The AI Software Market To $14 Trillion By 2030." *Ark Invest.* June 22. Accessed November 25, 2022. http://ark-invest.com/articles/analyst-research/productivity-gains-could-propel-the-ai-software-market-to-14-trillion-by-2030/.

Sundar, S. Shyam, and Clifford Nass. 2000. "Source Orientation in Human-Computer Interaction: Programmer, Networker, or Independent Social Actor." *Communication Research* 27 (6): 683–703. Accessed December 2022. doi:10.1177/009365000027006001.

Sutton, Zach. 2022. "Is AI Art Generation Going to Destroy Art as We Know It?" *Lensrentals.* October 3. Accessed November 17, 2022. https://www.lensrentals.com/blog/2022/10/is-ai-art-generation-going-to-destroy-art-as-we-know-it/.

Swinton, Jonathan. 2022. *Alan Turing's Manchester.* Gloucestershire: The History Press.

Talton, Jon. 2022. "Will Seattle ever reclaim its status as a vibrant 'superstar city'?." *Seattle Times.* December 30. Accessed January 2, 2023. https://www.seattletimes.com/pacific-nw-magazine/will-seattle-ever-reclaim-its-status-as-a-vibrant-superstar-city/.

Tangcay, Jazz. 2023. "'Hollywood 2.0': How the Rise of AI Tools Like Runway Are Changing Filmmaking." *Variety*, February 22: 1-11. https://variety.com/2023/artisans/news/artificial-intelligence-runway-everything-everywhere-all-at-once-1235532322/.

Tarantola, Andrew. 2017. "This is what AI sees and hears when it watches 'The Joy of Painting'." *Endgadget.* April 6. Accessed March 24, 2023. https://www.engadget.com/2017-04-06-ai-on-joy-of-painting-on-lsd.html.

The Daily Telegraph. 1946. "'ACE' will speed jet flying." *The Daily Telegraph*, November 8.

The London Times. 1949. "Calculus to Sonnet." *The London Times*, June 11: 2.

Theunissen, Yannick. 2022. "A.I. Tools Video Editors NEED to Start Using!" *Cinecom.net.* Diepenbeek: YouTube, December 16. Accessed December 2022. https://youtu.be/BhhFUGjyUNQ.

Thompson, Alan D. 2022. *A Comprehensive Analysis of Datasets Used to Train GPT-1, GPT-2, GPT-3, GPT-NeoX-20B, Megatron-11B, MT-NLG, and Gopher.* Research findings, San Francisco: Life Architect, 1-26. https://lifearchitect.ai/whats-in-my-ai-paper/.

—. 2022. "Devoxx Keynote - Artificial Intelligence: You Are Here." *Life Architect.* October 13. Accessed November 25, 2022. https://youtu.be/D3tTsoX02d8.

—. 2022b. "Integrated AI: The sky is infinite." *Life Architect.* YouTube, December 5. Accessed December 5, 2022. https://youtu.be/X8i9Op2HSjA.

—. 2022. "Transformative AI vs AGI vs Singularity." *Life Architect.* August 30. Accessed November 25, 2022. https://youtu.be/3cPIdSfM2I0.

Thompson, Clive. 2019. "What Will Happen When Machines Write Songs Just as Well as Your Favorite Musician?" *Mother Jones*, March/April: 7. Accessed December 10, 2022. https://www.motherjones.com/media/2019/03/what-will-happen-when-machines-write-songs-just-as-well-as-your-favorite-musician/.

Thoutt, Zack. 2017. *What are Creative Adversarial Networks (CANs)?* Analysis, Hackernoon, 1-17. https://hackernoon.com/what-are-creative-adversarial-networks-cans-bb81d09aa235.

Tillman, Maggie. 2022. "Amazon Go and Amazon Fresh: How the 'Just walk out' tech works." *Pocket Lint.* March 11. Accessed January 24, 2023. https://www.pocket-lint.com/gadgets/news/amazon/139650-what-is-amazon-go-where-is-it-and-how-does-it-work/.

Tufekci, Zeynep. 2022. "What Would Plato Say About ChatGPT?" *The New York Times*, December 15. Accessed December 16, 2022. https://www.nytimes.com/2022/12/15/opinion/chatgpt-education-ai-technology.html.

Turing, A. M. 1950. "Computing Machinery and Intelligence." *Mind* LIX (236): 433–460. Accessed December 2022. doi:10.1093/mind/LIX.236.433.

Turing, Alan M. 1950. "Computing Machinery and Intelligence." *Mind* (Oxford University Press) 59 (236): 433-460. Accessed December 2022. https://www.jstor.org/stable/2251299.

Turing, Alan M. 1996. "Intelligent Machinery, A Heretical Theory." *Philosophia Mathematica* 4 (3): 256-260.

Turing, Alan. 1936. "On Computable Numbers, with an Application to the Entscheidungsproblem." *Journal of Math* 58 (5): 239 - 265. doi:140.177.205.52.

Turkle, Sherry. 2005. *The Second Self, Twentieth Anniversary Edition: Computers and the Human Spirit.* Cambridge, MA: MIT Press. https://www.dhi.ac.uk/san/waysofbeing/data/communities-murphy-turkle-2005.pdf.

TwitchTracker. 2023. *Twitch Viewers.* April 12. Accessed April 12, 2023. https://twitchtracker.com/statistics/viewers.

Valéry, Paul. 1928. "La conquête de l'ubiquité." *Œuvres, tome II, Pièces sur l'art, Nrf, Gallimard, Bibl. de la Pléiade,* 1283-1287. http://cts.ulusofona.pt/wp-content/uploads/sites/93/2016/07/valery_conquete_ubiquite.pdf.

van Dijck, José. 2013. *The Culture of Connectivity: A Critical History of Social Media.* New York, NY: Oxford University Press. Accessed September 30, 2021. https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199970773.001.0001/acprof-9780199970773-chapter-1?print.

Vandeput, Jordy. 2022. "A.I. Tools Video Editors NEED to Start Using!" *Cinecom.net.* Diepenbeek: YouTube, December 16. https://youtu.be/BhhFUGjyUNQ.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. *Attention Is All You Need.* Research findings, arXiv:1706.03762. doi:10.48550/arXiv.1706.03762.

Verma, Pranshu. 2023. "They thought loved ones were calling for help. It was an AI scam." *The Washington Post.* March 5. Accessed March 13, 2023. https://www.washingtonpost.com/technology/2023/03/05/ai-voice-scam/.

Villegas, Ruben, Mohammad Babaeizadeh, Pieter-Jan Kindermans, Hernan Moraldo, Han Zhang, Mohammad Taghi Saffar, Santiago Castro, Julius Kunze, and Dumitru Erhan. 2022. *Phenaki: Variable Length Video Generation From Open Domain Textual Description.* Research findings, arXiv:2210.02399, 1-17. doi:10.48550/arXiv.2210.02399.

Vincent, James. 2022. "The scary truth about AI copyright is nobody knows what will happen next." *The verge.* November 15. Accessed February 4, 2023. https://www.theverge.com/23444685/generative-ai-copyright-infringement-legal-fair-use-training-data.

Vinge, Vernor. 1993. "The coming technological singularity: How to survive in the post-human era." In *Science fiction criticism: An anthology of essential writings*, 352-363. Washington, DC: NASA.

Waller, David. 2016. *Iron Men: How One London Factory Powered the Industrial Revolution and Shaped the Modern World.* London: Anthem Press.

—. 2017. "Why we should never listen to the Luddites." *CapX.* March 2. Accessed Janaury 16, 2023. https://capx.co/why-we-should-never-listen-to-the-luddites/.

Wang, Chengyi, Sanyuan Chen, Yu Wu, Ziqiang Zhang, Long Zhou, Shujie Liu, Zhuo Chen, et al. 2023. *Neural Codec Language Models are Zero-Shot Text to Speech Synthesizers.* Research findings, Microsoft Research, arXiv:2301.02111, 1-16. doi:10.48550/arXiv.2301.02111.

Wardle, Patricia. 1968. *Victorian lace (The Victorian collector series).* Traverse City, Michigan : Jenkins.

Waytz, Adam. 2019. *The Power of Human: How Our Shared Humnanity Can Help Us Create a Better World.* W.W. Norton & Company.

Webb, Aleksandra. 2022. "Artificial Intelligence and Digital Innovation in the Cultural Sectors: Technology, Sustainability, and new Creative Jobs. Are we Ready?" In *The Relevance of Artificial Intelligence in the Digital and Green Transformation of Regional and Local Labour Markets Across Europe: Perspectives on Employment, Training, Placement, and Social Inclusion*, edited by Christa Larsen, Jenny Kipper, Alfons Schmid and Marco Ricceri, 77-95. Baden-Baden: Nomos Verlagsgesellschaft. Accessed December 2022.

Webb, Michael. 2019. *The Impact of Artificial Intelligence on the Labor Market.* Research findings, Department of Economics, Stanford University, Paolo Alto: Stanford University - Department of Economic, 61. doi:10.2139/ssrn.3482150.

Weiner, Robert Paul. 2000. *Creativity and Beyond: Cultures, Values and Change.* Albany, NY: State University of New York Press. Accessed December 2022.

Wellner, Galit. 2021. "Digital Imagination, Fantasy, AI Art." *Foundations of Science* 27: 1445–1451. Accessed December 2022. doi:10.1007/s10699-020-09747-0.

Whiddington, Richard. 2022. "The French Collective Obvious Has Become One of the First A.I. Artists to Receive Gallery Representation." *The French Collective Obvious Has Become One of the First A.I. Artists to Receive Gallery Representation.* November 1. Accessed Janaury 6, 2023. https://news.artnet.com/market/french-collective-obvious-becomes-one-of-the-first-a-i-artists-to-receive-gallery-representation-2202005.

Wiggers, Kyle. 2022. "Microsoft brings DALL·E 2 to the masses with Designer and Image Creator." *TechCrunch.* October 12. Accessed November 21, 2022. https://techcrunch.com/2022/10/12/microsoft-brings-dall-e-2-to-the-masses-with-designer-and-image-creator/.

—. 2022. "Stability AI, the startup behind Stable Diffusion, raises $101M." *TechCrunch.* October 17. Accessed Janaury 17, 2023. https://techcrunch.com/2022/10/17/stability-ai-the-startup-behind-stable-diffusion-raises-101m/.

Wilson, H. James, Paul R. Daugherty, and Nicola Morini-Bianzino. 2017. "The Jobs That Artificial Intelligence Will Create." *MIT Sloan Management Review* 58 (4): 14-16. Accessed December 12 2022. http://mitsmr.com/2odREFJ.

Winograd, Terry. 1972. "Understanding Natural Language." *Cognitive Psychology* 3 (1): 1-191. doi:10.1016/0010-0285(72)90002-3.

Woodall, Tatyana. 2023. "Scientists Made a Mind-Bending Discovery About How AI Actually Works." *Vice Motherboard.* February 10. Accessed February 16, 2023. https://www.vice.com/en/article/4axjnm/scientists-made-discovery-about-how-ai-actually-works.

World Cube Association. 2023. *Results.* March 16. Accessed March 16, 2023. https://www.worldcubeassociation.org/results/rankings/333/average?&show=100+Results.

World Economic Forum. 2020. "The Future of Jobs Report 2020." *World Economic Forum Reports.* October 20. Accessed February 14, 2023. https://www.weforum.org/reports/the-future-of-jobs-report-2020/.

Write Brothers, Inc. 2022. *Dramatica: the next chapter in story development.* Accessed December 30, 2022. https://dramatica.com/.

Writers Guild of America, East. 2023. *Twitter.* March 22. Accessed April 7, 2023. https://bit.ly/42z2F7N.

Wu, Qiong, and Chunyan Miao. 2013. "Curiosity: From Psychology to Computation." *ACM Computing Surveys* 46 (2): 18:1-18:27. doi:10.1145/2543581.2543585.

Wundt, Wilhelm. 1874. *Grundzüge der Physiologischen Psychologie.* Leipzig: W. Engelman. https://pure.mpg.de/pubman/item/item_2407685_2/component/file_2407684/Wundt_1874_Grundzuege.pdf.

Xiang, Chloe. 2022. "Artists Are Revolting Against AI Art on ArtStation." *Motherboard.* Vice. December 14. Accessed February 9, 2023. https://www.vice.com/en/article/ake9me/artists-are-revolt-against-ai-art-on-artstation.

—. 2022. "Scientists Increasingly Can't Explain How AI Works." *Vice Motherboard.* November 1. Accessed February 16, 2023. https://www.vice.com/en/article/y3pezm/scientists-increasingly-cant-explain-how-ai-works.

Yalalov, Damir. 2022. "Midjourney and DALL·E Artist Styles Dump with Examples: 130 Famous AI Painting Techniques." *Metaverse Post.* August 11. Accessed March 14, 2023. https://mpost.io/midjourney-and-dall-e-artist-styles-dump-with-examples-130-famous-ai-painting-techniques/.

Yampolskiy, Roman V. 2019. *Unexplainability and Incomprehensibility of Artificial Intelligence.* Research findings, arXiv:1907.03869, 1-14. doi:10.48550/arXiv.1907.03869.

Zhang, Daniel, Nestor Maslej, Erik Brynjolfsson, John Etchemendy, Terah Lyons, James Manyika, Helen Ngo, et al. 2022. *The AI Index 2022 Annual Report.* Research findings, Paolo Alto: Stanford Institute for Human-Centered AI. Accessed December 8, 2022. https://aiindex.stanford.edu/report/.

Zhao, Chenqiu, and Anup Basu. 2020. "Dynamic Deep Pixel Distribution Learning for Background Subtraction." *IEEE transactions on circuits and systems for video technology a publication of the Circuits and Systems Society* 30 (11): 4192-4206. doi:10.1109/TCSVT.2019.2951778.

Zhou, Chunting, Graham Neubig, Jiatao Gu, Mona Diab, Paco Guzman, Luke Zettlemoyer, and Marjan Ghazvininejad. 2021. *Detecting Hallucinated Content in Conditional Neural Sequence Generation v3.* Research findings, arXiv, 1-16. doi:10.48550/arXiv.2011.02593.

Zi, W., L. S. Ghoraie, and S. Prince. 2019. *Few-shot learning and meta-learning.* Tutorial, Borealis AI. https://www.borealisai.com/research-blogs/tutorial-2-few-shot-learning-and-meta-learning-i/.