



DET PSYKOLOGISKE FAKULTET

***Virker psykologisk terapi basert på kunstig intelligens
mot symptomer på angst og depresjon?
En systematisk litteraturgjennomgang.***

HOVEDOPPGAVE

Profesjonsstudiet i psykologi

Lillian Synstelién

Høst/2023

Veileder: Anita Lill Hansen

Institutt for samfunnspsykologi

Forord

Gjennom studieløpet har jeg tilegnet meg kunnskap om behandling av psykiske lidelser, og samtidig fått kjennskap til utfordringene som psykisk helsevern står overfor. Blant de temaene som har blitt drøftet gjennom studieløpet har mangel på behandlingstilgjengelighet og økt etterspørsel vært sentralt. Samtidig har bruken av kunstig intelligens fått et fotfeste innenfor mange bransjer i de siste årene, og dette har fanget min interesse. Jeg har derfor ønsket å undersøke om kunstig intelligens kan sørge for at flere, særlig de som ikke blir prioritert i dagens helsevesen, kan få tilgang til god og effektiv hjelp gjennom fullautomatiserte intervensjoner. Jeg har fordypet meg i intervensjoner basert på kunstig intelligens for å redusere symptomer på angst og depresjon. Å arbeide med denne tematikken har blitt opplevd som svært meningsfullt, og jeg har utviklet en nysgjerrighet for krysningspunktet mellom psykologi og teknologi. Jeg håper denne litteraturgjennomgangen kan sette søkelys på bruk av moderne teknologi for å løse dagens utfordringer innen psykisk helse, og motivere til videre forskning.

Jeg ønsker å takke min veileder for gode innspill gjennom skriveprosessen. Jeg ønsker også å takke min samboer for støtte og motivasjon i en tidvis krevende skriveprosess. Det har vært svært lærerikt og givende å arbeide med denne hovedoppgaven, og jeg avslutter studiet med nye erfaringer og kunnskap som jeg håper å dra nytte av i min videre arbeidskarriere.

Merknad: Oppgaven bryter retningslinjer for APA-formatering ved å inkludere innholdsfortegnelse og ved å ha tabeller og figurer i løpende tekst. Dette er begrunnet i at det gir bedre leservennlighet.

Sammendrag

Den økende byrden av angst- og depressive lidelser kombinert med begrenset tilgjengelighet av tradisjonelle helsetjenester, har ført til at kunstig intelligens (AI) stadig blir mer utbredt for behandling av mentale helseutfordringer. Denne systematiske litteraturgjennomgangen har som primært mål å undersøke effekten av internettbaserte intervensjoner basert på kunstig intelligens for reduksjon av angst- og depressive symptomer. Et sekundært mål er å forstå mekanismene bak virkningen, med antropomorfisme og psykologisk avstand som foreslåtte medierende faktorer. Gjennom systematiske søk i databasene PsycInfo, Embase og Web of Science i august 2023, ble seks randomiserte kontrollerte studier (RCTs) inkludert i gjennomgangen. Resultatene av den primære undersøkelsen indikerer at samtlige RCT-studier viser en positiv effekt på reduksjonen av depressive symptomer, og varierte resultater når det gjelder reduksjon av angstsymptomer. Når det gjelder det sekundære målet, peker funnene mot antropomorfisme og psykologisk avstand som sentrale faktorer for hvordan brukere forholder seg til AI. Selv om funnene tyder på at AI-intervensjoner har potensial til å være kostnadseffektive, brukervennlige og tilgjengelige, spesielt for å redusere milde til moderate depressive symptomer, understrekes begrensningene knyttet til metodologiske faktorer, risiko for systematisk skjevhet og manglende helhetlig forskningsgrunnlag. Forslag til fremtidig forskning inkluderer forbedringer i metodologisk kvalitet, vurdering av variabler som innhold og varighet, samt utforskning av langtidseffekter og årsaker til effekt blant ulike populasjoner.

Nøkkelord: kunstig intelligens, AI, chatbots, angst, depresjon

Abstract

The increasing burden of anxiety and depressive disorders, coupled with limited accessibility to traditional healthcare services, has led to the growing prevalence of artificial intelligence (AI) in addressing mental health challenges. This systematic literature review aims primarily to investigate the effectiveness of internet-based interventions based on artificial intelligence in reducing anxiety and depressive symptoms. The secondary objective is to understand the mechanisms behind the impact, with anthropomorphism and psychological distance proposed as mediating factors. Through systematic searches in the databases PsycInfo, Embase, and Web of Science in August 2023, six randomized controlled trials (RCTs) were included in the review. The results of the primary analysis indicate that all included studies show a positive effect on reducing depressive symptoms, and varied results regarding the reduction of anxiety symptoms. Concerning the secondary objective, findings suggest that anthropomorphism and psychological distance are key factors in how users engage with AI. Although the findings suggest that AI interventions have the potential to be cost-effective, user-friendly, and accessible, especially for reducing mild to moderate depressive symptoms, limitations related to methodological factors, the risk of systematic bias, and a lack of comprehensive research foundation are emphasized. Suggestions for future research include improvements in methodological quality, assessment of variables such as content and duration, as well as exploration of long-term effects and causes of effectiveness among different populations.

Keywords: artificial intelligence, AI, chatbots, anxiety, depression

Innholdsfortegnelse

Innledning	7
Angst- og depressive lidelser	9
Behandling av angst og depresjon	10
Psykologiske Chatbots og AI-plattformer	12
Potensiell risiko ved psykologisk AI	13
Antropomorfisme og psykologisk avstand	14
Kontekst	16
Litteraturgjennomgangens problemstilling	18
Metode	18
Preliminære undersøkelser	18
Metodesøk.....	19
Inklusjons-og eksklusjonskriterier	20
Populasjon.....	20
Studiedesign.....	20
Intervensjoner	21
Utfallsmål.....	21
Språk og publikasjonsår.....	21
Data	22
Effektstørrelser.....	23
Risiko for systematisk skjevhet	23
Studiedesign.....	23
Utvalgs-kriterier	24
Utvalgsbias.....	24
Seleksjonsskjevhet	24
Publiserings-skjevhet	24
Rapporteringsbias	25
Resultater	25
Karakteristika ved de psykologiske intervensjonene.....	28
Fitzpatrick et al. (2017).....	28
Fulmer et al. (2018).....	28
Klos et al. (2021).....	29
Liu et al. (2022).....	29
He et al. (2022).	30
Sadeh-Sharvit et al. (2023).	31
Kontrollbetingelser	31

Utfallsmål.....	32
Effekt på symptomreduksjon	33
Tilbakemeldinger, akseptabilitet og engasjement.....	35
Fitzpatrick et al. (2017).....	35
Fulmer et al. (2018).....	36
Klos et al. (2021).....	36
Liu et al. (2022).....	36
He et al. (2022).	37
Sadeh-Sarvit et al. (2023).	37
Frafallsrater	38
Diskusjon	39
Oppsummering av funn.....	39
Utfallsmål på symptomreduksjon	39
Antropomorfisme og psykologisk avstand	41
Frafallsrater	43
Terapeutisk allianse	44
Terapeutisk allianse, antropomorfisme og psykologisk avstand.	45
Standardiserte mål på terapeutisk allianse.	46
Valg av kontrollgrupper.....	46
Deltakerkarakteristika og symptomnivå hos deltakerne	48
Systemfeil og innholdsmessige begrensninger	49
Kan man egentlig blinde deltakere i slike studier?	50
Behandlingsmetode og oppfølging	51
Begrensninger og styrker ved denne litteraturgjennomgangen	52
Implikasjoner og forslag til videre forskning.....	54
Konklusjon	56
Referanser	58
Tabeller	
Tabell 1	25
Tabell 2	35
Figurer	
Figur 1.....	20

Teknologien som omhandler kunstig intelligens (AI) har de siste årene gjort betydelige fremskritt, og produkter og tjenester som bruker AI har fått fotfeste i samfunnet. Samtidig forekommer angst- og depressive lidelser hos en femtedel av befolkningen globalt (Steel et al., 2014). Behandlingstilgjengeligheten blir stadig mindre, og digital behandling har derfor blitt et viktig studieområde som svar på den økte etterspørselen etter mental helsehjelp (Xiang et al., 2020) kombinert med mangelen på tilgang til tradisjonelle helsetjenester (Lattie et al., 2019; Xu et al., 2020). Mentale lidelser er svært utbredt og forblir ofte ubehandlet, særlig hos individer med milde til moderate symptomer. Mange begrensninger ved tradisjonell behandling kan potensielt overvinnes gjennom internettbaserte og AI-baserte intervensjoner (Ebert et al., 2019). Det globale markedet for kunstig intelligens hadde en verdi på 428 milliarder dollar i 2022 og forventes å vokse over 2000 milliarder dollar innen 2030 (Fortune Business Insights, 2023). Potensialet til AI-teknologi er stort fordi det kan implementeres i en rekke bransjer, som utdanning, helsevesen, hjemmeassistanse og handel (Li & Sung, 2021). Å tilby digital behandling av subterskel stemningslidelser vil kunne bidra til å forebygge risiko for videre utvikling av sykdomsforløp, og til å øke behandlingstilgjengeligheten. Fullautomatisert digital behandling kan være en kostnadseffektiv og engasjerende løsning som samtidig opprettholder positive resultater som kommer ved tidlig intervensjon (D'Alfonso et al., 2017).

Bruk av fullautomatiserte chatbots og AI-plattformer er et relativt nytt fenomen innen psykisk helse. Digitale behandlingsmetoder har hittil vært avhengige av menneskelige moderatorer for å levere terapeutisk innhold. Forskningsgrunnlaget knyttet til kunstig intelligens og psykisk helse er fremdeles begrenset, på tross av økende utbredelse de siste årene. I 2023 finnes det imidlertid enkelte chatbots og andre AI-verktøy som undersøker aspekter ved depresjon, angst, generelt velvære, livsstress, trivsel, avhengighet, rusbruk og autisme (Ahmed et al., 2021; Greer et al., 2019; Inkster et al., 2018; Mujeeb et al., 2017;

Prochaska et al., 2021; Tanaka et al., 2022). Flere av disse er undersøkt gjennom eksperimentelle forsøk for å teste effektivitet, akseptabilitet og brukervennlighet (Abd-alrazaq et al., 2019; Ali et al., 2020; Yasavur et al., 2014). En tidlig studie utviklet en chatbot som var ment å øke fysisk aktivitet og demonstrerte dessuten at det kan etableres et terapeutisk forhold mellom en klient og en chatbot (Bickmore et al., 2005).

Siden bruken av AI innen psykisk helse stadig blir mer utbredt og har potensial til å overvinne behandlingsbarrierer, samtidig som byrden av angst og depresjon globalt er stor, vil det være nødvendig å undersøke i hvilken grad fullautomatiserte intervensjoner basert på kunstig intelligens kan bidra til symptomreduksjon ved angst og depresjon. Det vil også være relevant å undersøke hvilke faktorer som ligger bak den potensielle symptomreduksjonen. Det trengs mer kunnskap om hvorvidt disse intervensjonene har effekt, i hvilken grad de virker og hvordan de virker. For å undersøke dette vil denne systematiske litteraturgjennomgangen undersøke de eksisterende RCT-studier som undersøker effekten av fullautomatiserte chatbots eller AI-plattformer mot symptomer på angst og depresjon.

Angst- og depressive lidelser

De vanligste symptomene på depresjon ifølge DSM-5 er vedvarende tristhet, tap av interesse og glede, endringer i vekt eller appetitt, søvnproblemer, tretthet eller energimangel, følelse av verdiløshet eller skyld, konsentrasjonsproblemer og selvmordstanker eller selvmordsforsøk (APA, 2013). Symptomer på angstlidelser kan variere ut ifra den aktuelle form for angstlidelse, og inkluderer overveldende bekymring eller frykt, fysiske symptomer som muskelspenning, skjelvinger, rastløshet, svette, hjertebank eller kortpustethet, kognitive symptomer som problemer med konsentrasjon, forvirring, frykt for å miste kontrollen eller gå fra vettet og unngåelse, søvnproblemer, frykt for panikkanfall, muskelspenninger, irritabilitet og frykt for sosiale situasjoner (APA, 2013).

Depresjon og angstlidelser har i lang tid utgjort betydelige komponenter av den globale sykdomsbyrden. I henhold til Global Burden of Disease (GBD) 2019-studien rangerte depresjon og angst blant de 25 fremste årsakene til sykdomsbyrde på verdensbasis i 2019 (Abrams et al., 2020). I tillegg har COVID-19-pandemien skapt en kontekst der mange determinanter for mental helse har blitt påvirket. Faktorer som sosiale begrensninger, nedstengninger av sentrale arenaer, tap av levebrød og endringer i økonomisk aktivitet har potensial til å påvirke befolkningens mentale helse i stor grad og over lang tid. Rater av angst og depresjon blant voksne i USA var omtrent 4 ganger høyere mellom april 2020 og august 2021 enn de var i 2019 (DeAngelis, 2021). Santomauro et al. (2021) kvantifiserte i sin studie innflytelsen av COVID-19 pandemien på forekomsten og byrden av depressiv lidelse og angstlidelser globalt. Studien fant at det ble anslått en økning av depressiv lidelse på 53,2 millioner (en økning på 27,6%) tilfeller globalt og 76,2 millioner tilfeller av angstlidelser (en økning på 25,6%) globalt. Samlet forårsaket depressiv lidelse 49,4 millioner -tapte leveår med helsetap (DALYs), mens angstlidelser forårsaket 44,5 millioner DALYs globalt i 2020. Begrensninger i behandlingstilgjengelighet og akseptabilitet ble tydeligere med økende risiko for mentale helseproblemer på grunn COVID-19-pandemien (Wu et al., 2021). Disse funnene gir grunnlag for videre diskusjon om tilgang til behandling og bedre måter å tilby omsorg og veiledning til de som opplever symptomer på angst og depresjon.

Behandling av angst og depresjon. Når det gjelder stemningslidelser som angst og depresjon, anerkjennes kognitiv atferdsterapi (CBT) som en bredt akseptert og effektiv evidensbasert form for psykoterapi. CBT fungerer som en av de sentrale teoretiske rammene for digitale intervensjoner, inkludert de som er basert på kunstig intelligens (Butler et al., 2006). Tradisjonell behandling ansikt-til-ansikt er fremdeles ansett som standard omsorg for de med klinisk depresjon og angst, men foreløpige studier antyder at selvbetjente datamaskinbaserte kognitive og atferdsmessige intervensjoner gir lignende resultater og er effektive i

behandlingen av subterskelstemningslidelser (Eells et al., 2014). Videre viste en studie at nesten 70% av pasienter uttrykte interesse for å bruke mobile helseapper for å overvåke og administrere sin mentale helse selv (Torous et al., 2014).

Økt prevalens av stemningslidelser og begrenset tilgjengelighet og akseptabilitet av tradisjonelle psykoterapier har økt utfordringene knyttet til behov for psykisk helsehjelp. Barrierefaktorer som frykt for stigmatisering, praktiske hindringer, kostbare behandlinger, begrenset tilgang til erfarne terapeuter og lange ventetider har vist seg å begrense tilgangen til mentale helsetjenester (Christensen, 2007; Clement et al., 2015; Kazdin & Blase, 2011). Det er ifølge Oladeji og Gureje (2016) omtrent 9 psykiatere per 100 000 mennesker i utviklede land, og så få som 0,1 per 1 000 000 i lavinntektsland. Denne utilstrekkeligheten i å imøtekomme nåværende eller fremtidig etterspørsel etter omsorg har ført til forslaget om teknologi som en løsning.

Behandling av mentale helseproblemer har gjennomgått en digital transformasjon gjennom datamaskiner, internett, mobile enheter, mobile programvareapper og virtuell virkelighet (Fairburn & Patel, 2017). Internettbaserte intervensjoner kan fungere som en effektiv måte å nå mennesker på, og kan bidra til å redusere behandlingsbarrierer. Internettbaserte psykologiske intervensjoner har vist seg å være effektive fordi brukere kan ha private, selvstyrte intervensjoner til enhver tid og uten geografiske begrensninger, og reduserer stigmatiseringen knyttet til å oppsøke en terapeut (Andersson, 2016; Kaiser et al., 2021). Tidlige bevis antyder at pasienter åpner seg mer når de bruker en mental helse-applikasjon enn under ansikt-til-ansikt terapi (Rost et al., 2017). Med en app for pasienter som led av selvmordstanker, rapporterte flere deltakere selvmordstanker ved hjelp av appen enn de gjorde med den tradisjonelt administrerte Pasient Health Questionnaire (PHQ-9) (Torous et al., 2015). En annen studie fant at individer kan danne et positivt bånd med en AI-

chatbot på grunn av dets personlighetstrekk, som å være omsorgsfull, lydhør og ikke-dømmende (Dosovitsky & Bunge, 2021).

Psykologiske Chatbots og AI-plattformer

En chatbot eller AI-plattform er en digital teknologi som simulerer samtaler med brukere via tekst, stemme, digitale mennesker eller virtuelle mennesker, med hjelp av programvare basert på kunstig intelligens eller maskinlæring (McTear et al., 2016). Kunstig intelligens refererer til "en gren av datavitenskap som handler om simuleringen av intelligensatferd hos datamaskiner" eller "evnen til en maskin til å etterligne intelligent menneskelig atferd" (Merriam-Webster, u.å). Chatbots kan etterligne menneskelig atferd og tilby et oppgaveorientert rammeverk med utviklende dialog. Enkelte chatbots oppfatter og benytter emosjoner og ansiktsuttrykk under interaksjoner med individer i sanntid (Ma et al., 2019). Psykologiske chatbots er designet til å bruke interaktive funksjoner, og ikke bare psykologisk opplæring for å lette på psykoterapiprosessen (Laranjo et al., 2018). De fleste chatbots kan gi tjenester uavhengig av andre faktorer, og uten behov for deltakelse og veiledning fra menneskelige terapeuter (Gaffney et al., 2019). Kunstig intelligens har kapasitet til å håndtere flere datapunkter samtidig, er immun mot utbrenthet, kan anvende logikk, lage prediksjoner basert på erfaring eller input, og kan fungere som et verdifullt verktøy for klinikere. AI kan støtte tolkninger av behandlingsdata og reduserer den administrative byrden (Antoniadi et al., 2021; Kellogg & Sadeh-Sharvit, 2022). Chatbots gir også bedre tilgjengelighet enn tradisjonell ansikt-til-ansikt terapi, og brukere kan oppleve følelsen av å bli ledsaget og forstått (Richards & Richardson, 2012).

Det er økt behov for, og fremvekst av, digitaliserte intervensjoner med fokus på bedring av psykisk helse. Det er imidlertid få av disse som er fullautomatiserte. I de siste årene har det dukket opp flere psykologiske chatbots som er konstruert for å forbedre psykisk helse, og deres effektivitet er testet gjennom randomiserte kontrollerte forsøk for ulike mentale

helseutfordringer. Av de som er direkte undersøkt mot angst og depresjon finnes Woebot (Fitzpatrick et al., 2017), Tess (Fulmer et al., 2018; Klos et al., 2021), XiaoNan (Liu et al., 2022) og XiaoE (He et al., 2022). Det finnes også AI-drevne plattformer som brukes som hjelpemiddel av behandlere i behandling av depresjon og angst basert på kunstig intelligens, og som er testet gjennom RCT (Sadeh-Sharvit et al., 2023).

Potensiell risiko ved psykologisk AI. Studier har vist at det finnes flere risikoer ved å benytte seg av psykologiske chatbots. Dette inkluderer misforståelser som kan føre til ineffektive eller til og med skadelige inngrep, manglende varslingsystemer for kriser og mangel på personvernbeskyttelse (Xu & Zhuang, 2022). Integreringen av kunstig intelligens i helsevesenet vekker bekymringer knyttet til personvern, inkludert overvåking av data og mulig misbruk av informasjon. En annen bekymring rundt bruken av AI innen psykisk helse er at det er et gap mellom de hurtige teknologiske fremskrittene og implementeringen av disse verktøyene i kliniske settinger. Dessuten er det blitt påvist at intervensjonene ofte utvikles uten standardiserte rutiner for etiske vurderinger (Ienca et al., 2018). Det finnes i tillegg utfordringer knyttet til bruk av AI når det gjelder differensialdiagnostikk og å støtte kliniske beslutninger, i likhet med når menneskelige agenter diagnostiserer psykisk lidelse (Vermani et al., 2011). Psykiske lidelser, i motsetning til medisinske, er ofte heterogene og komplekse, og diagnostiske kriterier inkluderer subjektive indikatorer, variasjoner i symptompresentasjon, komorbiditet og en dynamisk natur av symptomer (Yan et al., 2023), som gjør det utfordrende å stille riktig diagnose.

Fiske og kolleger (2019) reiser flere bekymringer knyttet til bruken av AI innen psykisk helse. For det første peker de på etiske og sosiale bekymringer knyttet til mangel på grundige forskningsstudier og manglende etiske retningslinjer for bruk av AI i kliniske settinger. Videre er det mangel på klare retningslinjer for utvikling, klinisk integrasjon og opplæring av AI-applikasjoner innen psykisk helse, noe som kan true pasientsikkerheten. Det er også en

bekymring for at integreringen av AI kan bidra til å redusere tilbudet av tradisjonelle tjenester, med risiko for færre helseressurser og økt ulikhet. Pasienters autonomi kan utfordres når det gjelder forståelse og samtykke til bruk av AI, spesielt for sårbare grupper. Langsiktige virkninger på pasienter og samfunnet er usikre, inkludert avhengighet av AI-applikasjoner og endringer i sosiale verdier. Det er også en frykt for at mennesker kan utvikle en overdreven avhengighet av AI for terapeutiske formål, da tilgangen ikke er begrenset på samme måte som med menneskelige klinikere. Bekymringer om "teknologiavhengighet" fra spill og sosiale medier kan også gjelde for forholdet til AI, med frykt for at AI kan erstatte, snarere enn supplere, personlig helsetjeneste (Fiske et al., 2019).

Antropomorfisme og psykologisk avstand

I lys av den relativt nylige implementeringen av kunstig intelligens innen psykisk helse, mangler det et solid teoretisk rammeverk for å forstå samspillet mellom mennesker og AI-agenter. I denne litteraturgjennomgangen er det et sekundært spørsmål å utforske hvilke faktorer som påvirker effekten av slike intervensjoner. Med AI-agenter som har menneskelignende egenskaper, oppstår spørsmål om hvordan disse trekkene påvirker samspillet med brukerne. Innenfor CASA-paradigmet (Computers are Social Actors Paradigm) har tidligere forskning fastslått at individer bruker sosiale normer og reagerer sosialt når de samhandler med datamaskiner (Reeves & Nass, 1996). Dette har ført til omfattende studier av begrepene antropomorfisme og psykologisk avstand med hensyn til menneske-AI-interaksjon (Li & Sung, 2021).

Teorien om antropomorfisme hevder at mennesker har en tendens til å tillegge ikke-menneskelige agenter menneskelignende egenskaper. Dette påvirker kognitive, holdningsmessige og atferdsmessige resultater av interaksjonen, og er en utbredt måte å tenke på (Epley et al., 2007; Mithen & Boyer, 1996). Psykologisk avstand refererer ifølge konstrualitetsteorien til subjektiv oppfatning av avstander i tid, rom og sosiale forhold til et

mål. Denne oppfatningen påvirker vurderinger, evalueringer, prediksjoner, valg og atferd (Lieberman & Trope, 1998; Trope & Liberman, 2010). Psykologisk avstand er ikke alltid objektiv og påvirkes av faktorer som valens og intensitet av følelser og handlinger, ønskelighet, oppfattet kontroll og motivasjon (Balci et al., 2010; Han et al., 2018; Van Boven et al., 2010). Særlig dimensjonen sosial avstand har blitt undersøkt som medierende faktor i antropomorfisering av AI-assistenter, selv om alle dimensjoner av psykologisk avstand er korrelerte (Trope & Liberman, 2010).

Flere studier har undersøkt antropomorfisme og psykologisk avstand under eksperimentelle forhold. Li og Sung (2021) viste i sin studie at når brukere tilskriver AI-assistenter menneskelige egenskaper, blir deres vurdering mer av assistenten mer positiv. Denne sammenhengen mellom antropomorfisme og vurdering ble mediert av redusert psykologisk avstand. En annen studie av Park et al. (2023) utforsket hvordan menneskelig representasjon i chatbots påvirket brukeres intensjon om å følge helseanbefalinger, og viste at chatbots med menneskelige representasjoner økte intensjonen til å følge anbefalinger. Denne effekten ble også mediert av psykologisk avstand, særlig oppfattet nærhet og tillit til chatboten. Chung og Han (2022) fant i sin studie at bruk av emojis for å representere chatbots reduserte oppfattet psykologisk avstand. Psykologisk avstand har også blitt knyttet til flere positive interaksjonsresultater som generell positivitet, holdninger, tiltrekning, dømmekraft og ettergivenhet med chatbot-anbefalinger (Ahn et al., 2021; Goethals & Nelson, 1973; Nowak & Rauh, 2005; Trope & Liberman, 2010).

Viktigheten av antropomorfisme og psykologisk avstand i sammenhengen med menneske-AI-interaksjon ligger i det faktum at det gir en teoretisk basis for hvordan mennesker samhandler med en ikke-menneskelig enhet som AI, og belyser de psykologiske mekanismene bak menneske-AI-interaksjon (Duffy, 2003). Denne forskningen gir innsikt i hvorfor brukere foretrekker og har en mer positiv holdning til AI-enheter som de oppfatter

som mer menneskelige. Studier som støtter dette teoretiske grunnlaget for menneske-robot-interaksjon har funnet at antropomorfe agenter oppfattes som mer responsive, tiltrekkende, troverdige og oppfattes som mer positive (Go & Sundar, 2019; Li & Sung, 2021; Nowak & Rauh, 2005).

Det finnes forskning som understreker negative sider ved menneskelignende trekk hos AI-agenter, eller som stiller seg kritisk til antropomorfisme som teoretisk grunnlag for menneske-AI-interaksjon. Et perspektiv er at menneskelignende egenskaper kan øke brukernes forventninger til interaksjonen, som kan føre til negative utfall dersom interaksjonskvaliteten ikke møter forventningene (Go & Sundar, 2019). Det er også påpekt at den høye oppfattede menneskeligheten kan ha negative virkninger knyttet til personvern hensyn (Miner et al., 2017), og at grad av likhet mellom AI-agenten og brukeren kan være avgjørende for om man foretrekker AI-enheten, uavhengig av grad av menneskelighet (Nowak & Rauh, 2005). Reeves og Nass (1996) påpeker også at mennesker antropomorfiserer alt de møter i en viss grad, også objekter, selv om de ikke er menneskelige. Nowak og Rauh (2005) fant i sin studie at det heller var grad av androgynitet enn antropomorfisme som bestemte grad av troverdighet eller tiltrekningskraft hos en AI-enhet. Selv om noen studier peker på negative sider, og det ikke er etablert en tydelig enighet om hvilke begreper som danner det teoretiske rammeverket rundt menneske-AI-interaksjon, kan antropomorfisme og psykologisk avstand som et teoretisk rammeverk bidra til å forstå kompleksiteten i menneske-AI-interaksjon.

Kontekst

Det er nødvendig å undersøke effekten av psykologisk terapi basert på kunstig intelligens (AI) på symptomer relatert til angst og depresjon av flere grunner. For det første utgjør den økende utbredelsen av angst- og depresjonslidelser på globalt nivå en betydelig byrde for samfunnet, og tradisjonelle behandlingsmetoder har vist seg utilstrekkelige for å

møte den økende etterspørselen etter mental helsehjelp. I tillegg har teknologiske fremskritt og digitalisering av helsevesenet ført til økende interesse for AI som en potensiell løsning. Den raske veksten i det globale markedet for AI og dens anvendelser i en rekke bransjer, inkludert helsevesenet, understreker behovet for å utforske potensialet til AI i behandling av angst og depresjon. Videre, med begrensningene knyttet til tilgjengeligheten av tradisjonelle helsetjenester, er det nødvendig å utforske effektiviteten av AI-løsninger for å tilby bedre tilgjengelig mental helsehjelp. Dette er spesielt viktig i en tid der økonomiske, sosiale og helsemessige faktorer som COVID-19-pandemien har forsterket behovet for slike tjenester.

Den økende bruken av psykologiske chatbots og AI-plattformer rettet mot ulike mentale helseutfordringer gir en indikasjon på AI-teknologiens potensial. Flere av disse løsningene har blitt testet gjennom eksperimentelle forsøk for å evaluere deres effektivitet og akseptabilitet. Som sådan, er det nødvendig å gjennomføre en systematisk litteraturgjennomgang for å vurdere i hvilken grad fullautomatiserte intervensjoner basert på kunstig intelligens kan bidra til symptomreduksjon ved angst og depresjon. En gjennomgang som undersøker dette systematisk, eksisterer ikke ut ifra undertegnede (LS) sin kunnskap. En slik gjennomgang vil kunne bidra til å gi en oversikt over forskningsfeltet og potensielle styrker og svakheter ved bruk av AI for å behandle symptomer på angst og depresjon. Denne gjennomgangen vil kunne bidra til å øke kunnskapsgrunnlaget på området, men også hjelpe med å belyse risikoer og utfordringer som følger med bruken av AI i psykisk helse, i tillegg til å belyse hvordan brukere oppfatter og samhandler med AI-assistenter. Slik kunnskap er avgjørende for å utvikle retningslinjer og forbedre tilgjengeligheten av effektive mentale helsetjenester i en tid hvor behovet er stort, og hvor teknologiske fremskritt gir nye muligheter innen behandling av angst og depresjon. Oppgaven forsøker samtidig å demonstrere sentrale forskningsprinsipper undertegnede (LS) har tilegnet seg i løpet av studiet.

Litteraturgjennomgangens problemstilling

Det primære målet med denne litteraturgjennomgangen er å studere hvorvidt psykologisk terapi basert på kunstig intelligens virker mot symptomer på angst og depresjon. Problemstillingen besvares ved å undersøke de eksisterende RCT-studier som tester effekten av behandling basert på kunstig intelligens for å redusere symptomer på angst og depresjon. Symptomreduksjon undersøkes ved å se på grenseverdier for klinisk signifikante eller ikke-klinisk signifikante symptomer på angst og depresjon ved bruk av reliable og validerte kartleggingsverktøy. Effektiviteten av bruk av kunstig intelligens for å redusere symptomer på angst og depresjon vurderes også ved å undersøke karakteristika ved de inkluderte studiene, rapporterte effektstørrelser, frafallsrater og tilbakemeldinger. Et sekundært mål i denne litteraturgjennomgangen er å undersøke hvordan behandling basert på kunstig intelligens har virkning. Mekanismene bak virkningen undersøkes gjennom å utforske frafallsrater, tilbakemeldinger, karakteristika ved design og innhold, brukeropplevelser og tilbakemeldinger, og vil diskuteres i sammenheng med teori om psykologisk avstand og antropomorfisme.

Metode

Preliminære undersøkelser

I forkant av litteratursøket ble det gjennomført usystematiske søk for å få oversikt over litteraturen på feltet. Disse ble gjennomført i august 2023, i databasene Google Scholar, Web of Science, Psycinfo og Embase. De usystematiske søkene gjorde undertegnede (LS) oppmerksom på at litteraturen på området var begrenset, både tidsmessig og i antall. Det ble også oppdaget at terminologien på fagfeltet var variert, hvor studier som undersøkte samme fenomener benyttet seg av forskjellige betegnelser for samme fenomen. Dette ble tatt høyde for i det systematiske litteratursøket. Med bakgrunn i disse funnene ble det planlagt for et litteratursøk med brede inklusjonskriterier, som skulle reflektere den varierte bruken av

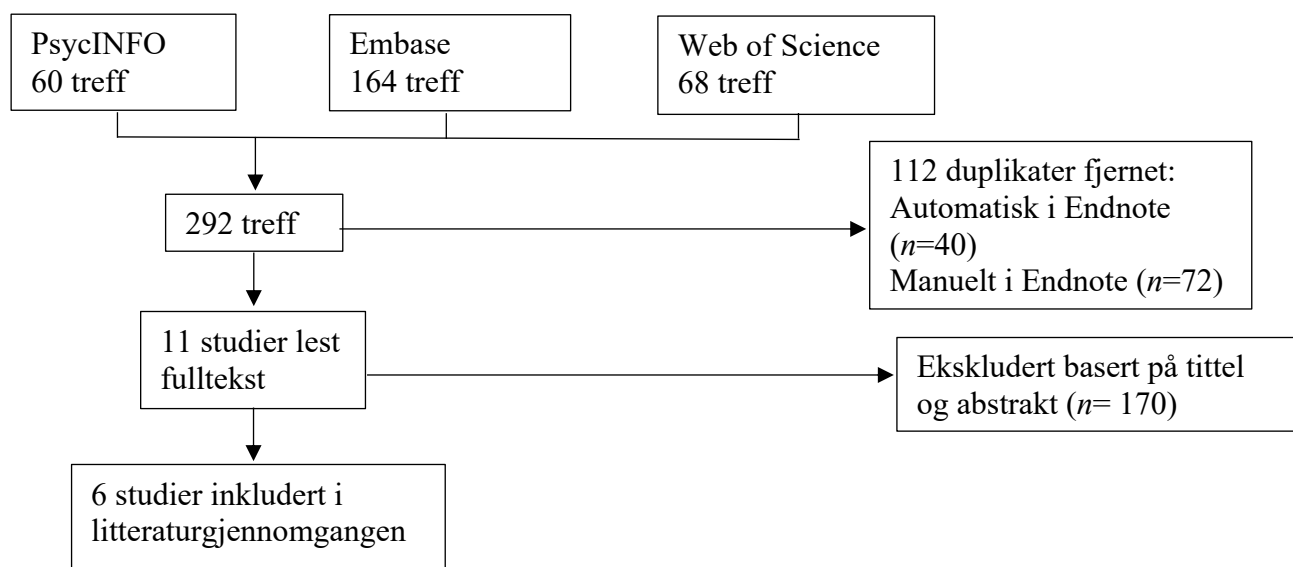
terminologi på feltet, og for å fange opp alle relevante artikler. Siden fagområdet er i et tidlig stadium, ble det også bestemt at kun RCT-studier skulle inkluderes i gjennomgangen, for å sikre god forskningsmessig kvalitet på de inkluderte studiene, og for å kompensere for et mulig begrenset antall studier.

Metodesøk

Det ble gjennomført systematiske søk i PsycInfo, Embase og Web of Science. Databasesøkene ble gjennomført 23. august 2023 av undertegnede (LS). Litteratur ble identifisert gjennom følgende søkeord: ("Artificial intelligence" OR "chatbot*" OR "Therap* bot" OR "therap* bots" OR "machine learning" OR "Avatar therap*" OR "Conversational agent" OR "Automated psychotherap*" OR "Virtual therapeutic agent*" OR "AI-supported psych*") AND ("anxiety" OR "depression" OR "mood disorder*" OR "Affective disorder*") AND ("Treatment" or "therap*"). I de databasene hvor det ikke var mulig å filtrere på RCT studier ble også søkeordene "randomized" OR "RCT" benyttet, da det kun var ønskelig å undersøke RCT studier. I PsychInfo Og Embase ble det søkt etter litteratur i tittel, abstract og nøkkelord (ti, ab, id). I Web of Science ble det søkt etter topic.

Søk i PsycInfo, Embase og Web of Science genererte totalt 306 treff (henholdsvis 60 + 164 + 68). Av disse treffene ble duplikater fjernet (112) og inklusjons –og eksklusjonskriterier knyttet til type studier (RCT), tema og språk (engelsk og skandinaviske språk) ble anvendt. Dette resulterte i totalt 11 treff som ble screenet for tittel og abstrakt. Av disse ble samtlige studier lest til fulltekst, hvorav 6 RCT-studier oppfylte inklusjons og eksklusjonskriteriene. Det ble ikke sortert på årstall fordi det ikke eksisterer RCT-studier som har undersøkt bruk av kunstig intelligens for å redusere symptomer på angst og depresjon før 2017, ut ifra de preliminære undersøkelsene som ble gjennomført. Syv publiserte RCT-studier som undersøker effekten av kunstig intelligens for å redusere symptomer på angst og depresjon ble oppdaget. Én studie (Danieli et al., 2022) ble ekskludert grunn av mangelfullt

studiedesign. Totalt seks RCT-studier publisert mellom 2017 og 2023 ble inkludert i denne litteraturgjennomgangen. Litteratursøket ble avsluttet 29. august 2023.



Figur 1

Flytdiagram søkeprosess

Inklusjons-og eksklusjonskriterier

Populasjon. I denne oppgaven undersøkes mennesker over 17 som rapporterer om kliniske eller ikke-kliniske angst eller depresjonssymptomer. Det er ikke gjort ekskludering med bakgrunn i geografi, kjønn eller kliniske grenseverdier. Årsaken til at alderskriteriet er satt til 17 år er fordi én av studiene undersøker college-studenter som er minimum 17 år.

Studiedesign. Denne litteraturgjennomgangen gjennomgår de eksisterende RCT-studier som undersøker effekten av bruk av kunstig intelligens i form av chatbots eller AI-plattformer for å behandle angst og/eller depresjonssymptomer. Dette inkluderer studier som både undersøker bruk av chatbots eller AI plattformer benyttet av forsøksperson og som er benyttet av behandler. De inkluderte studiene må oppfylle kravene for en randomisert kontrollert studie ved å ha en tydelig randomisering, inklusjons- og eksklusjonskriterier og forhåndsdefinerte statistiske metoder for dataregistrering og analyse, slik at risiko for selektiv rapportering av resultater er minimert. De må også velge relevante og validerte utfallsmål. Studiene som er

inkludert i denne gjennomgangen må ha undersøkelse av effekt på angst og/eller depresjon som primært forskningsspørsmål. I tillegg er studier som ikke var publisert i et fagfelleurdert tidsskrift eller på engelsk ekskludert, det samme ble rene kvalitative studier, protokoller, avhandlinger og casestudier.

Intervensjoner. Studiene i litteraturgjennomgangen må inneholde intervensjoner som er gjennomført av ikke-menneskelig art som sin primære administrasjonsmetode i den eksperimentelle gruppen. Det vil si chatbots, avatarer, automatiserte apper eller andre plattformer basert på kunstig intelligens og gjennom internett-baserte løsninger. Intervensjoner, oppfølging eller tilleggsbehandling hos deltakere som foregår ved hjelp av tradisjonell ansikt-til-ansikt terapi kan inkluderes, men ekskluderes dersom det benyttes som primær administrasjonsmetode. Kontrollgruppe må tilbys placebo eller en anerkjent tilgjengelig alternativ behandlingsmetode.

Utfallsmål. Studiene som er med i denne litteraturgjennomgangen undersøker symptomatologi for angst og depresjon i henhold til kliniske grenseverdier på validerte selvrapporteringsmål. Dette inkluderer primært The Patient Health Questionnaire-9 Pasienthelseundersøkelsen (PHQ-9) for depresjon, og Generalized Anxiety Disorder 7 (GAD-7) for angstsymptomer. Andre kvalitative eller kvantitative mål kan inkluderes dersom de kan ses i sammenheng med effekt på symptomer, men kun dersom PHQ-9 eller GAD-7 er benyttet som primært utfallsmål. Studier som har som primært utfallsmål å undersøke spesifikke angstlidelser som panikk lidelse, tvangslidelse eller andre spesifikke angstlidelser, er ekskludert fra litteraturgjennomgangen. Studier som måler generelt velvære, livsstress eller andre ikke-symptomatologiske mål for depresjon er også ekskludert fra denne gjennomgangen.

Språk og publikasjonsår. Siden bruk av kunstig intelligens for behandling av psykisk helse er et relativt nytt forskingsområde, er det ikke funnet RCT-studier som undersøker effekten av kunstig intelligens for symptomer på angst og depresjon før 2017. Denne

litteraturgjennomgangen undersøker derfor studier som er gjennomført mellom 2017 og august 2023. RCT-studiene skal være skrevet på engelsk eller skandinaviske språk.

Følgende eksklusjonskriterier foreligger: 1) studier som ikke oppfyller randomiserte kontrollerte forsøkskriterier, 2) studier som mangler relevante og validerte utfallsmål (manglende bruk av PHQ-9, eller GAD-7 som primære utfallsmål), 3) studier som har spesifikke angstlidelser eller ikke-symptomatologiske mål som primært utfallsmål, 4) studier som ikke er publisert i fagfelleverderte tidsskrifter, 5) studier som ikke er skrevet på engelsk eller skandinaviske språk, 6) studier som er publisert etter august 2023, 7) studier som evaluerer teknologibaserte løsninger som ikke er gjennomført med bruk av kunstig intelligens, 8) studier som ikke bruker en psykologisk kunstig intelligens som primær intervensjon, 9) studier som undersøker populasjon under 17 år.

Data

Et datainnsamlings skjema (*Tabell 1*) ble utarbeidet for å hente ut relevant informasjon fra de seks selekterte RCT-studiene. Dataene inkluderer følgende:

- a. Førsteforfatter, publikasjonsår
- b. Formål ved studien
- c. Type studie
- d. Karakteristika ved deltakere – totalt antall deltakere, gjennomsnittlig alder/ aldersspenn og etnisitet/nasjonalitet;
- e. Inklusjons-og eksklusjonskriterier
- f. Frafall
- g. Intervensjonskarakteristika; type, administrasjonsmetode, antall sesjoner/interaksjoner, behandlingsvarighet og oppfølging
- h. Utfallsmål;
- i. Kontrollbetingelser

- j. Administratorer/behandlere;
- k. Studiens finansieringskilde(r)

Effektstørrelser

Effektstørrelser er kvantitative mål som brukes for å indikere styrken eller størrelsen på en effekt i en gitt studie eller eksperiment. Disse målene hjelper for å forstå den praktiske betydningen av resultatene og sammenligne effekter på tvers av forskjellige studier.

Signifikante resultater i denne systematiske litteraturgjennomgangen vil bli presentert i form av Cohen's d der de er oppgitt. Cohen's d er en statistisk effektstørrelse som brukes for å måle den standardiserte forskjellen mellom gjennomsnittene i to grupper. Den hjelper med å evaluere hvor stor den praktiske eller kliniske forskjellen er mellom gruppene ved å ta hensyn til variasjonen i dataene. En Cohen's d-verdi på 0.2 indikerer en liten effekt, 0.5 en moderat effekt, og 0.8 en stor effekt.

Risiko for systematisk skjevhet

I en systematisk litteraturgjennomgang kan systematiske feil i de inkluderte studiene føre til en skjevhet i resultatene, som igjen kan påvirke påliteligheten og gyldigheten av konklusjonene (Higgins et al., 2011). Dette kan redusere evnen til å gi en nøyaktig og pålitelig oppsummering av studiene som er inkludert i gjennomgangen. For å identifisere og vurdere risikoen for slike systematiske skjevheter i denne gjennomgangen, blir studiedesign, utvalgsriterier, utvalgsbias, seleksjonsskjevhet, publiseringsskjevhet og rapporteringsbias tatt høyde for i den grad det lar seg gjøre.

Studiedesign. Randomiserte kontrollerte studier (RCT-studier) er anerkjente som grundige forskningsdesign innen medisinsk og vitenskapelig forskning. RCT-studier er kjent for å minimere risikoen for systematiske skjevheter. Alle studiene som er inkludert i denne gjennomgangen, er derfor RCT-studier. Dette valget skyldes det relativt utforskede

forskningsfeltet som gjennomgangen undersøker, og viktigheten av å etablere årsakssammenhenger i en begrenset studiepopulasjon.

Utvalgs-kriterier. Veldefinerte utvalgs-kriterier er avgjørende for pålitelige resultater og reduserer potensialet for bias og feil. Disse kriteriene sikrer at studien måler nøyaktig det den er ment å måle, samtidig som de hjelper til med å minimere subjektivitet. Bare studier med tydelig definerte inklusjons- og eksklusjonskriterier blir inkludert i gjennomgangen, og de anses som relevante for dens formål. Likevel kan det oppstå usikkerhet knyttet til objektiviteten av disse kriteriene, inkludert valg av deltakere, studiekvalitet og måling av effekter.

Utvalgsbias. Utvalgsbias oppstår når studiens utvalg ikke er representativt for den populasjonen eller gruppen som forskeren ønsker å trekke konklusjoner om. Utvalgsbias kan føre til unøyaktige og feilaktige resultater og begrenset generaliserbarhet. Denne gjennomgangen fokuserer på en populasjon over 17 år med symptomer på angst og depresjon, og studier som involverer andre populasjoner eller symptomer, blir ekskludert.

Seleksjonsskjevhet. Dette refererer til skjevhet som oppstår når det er mangelfull eller partisk utvelgelse av deltakere i studien, noe som kan påvirke resultatenes generaliserbarhet og gyldighet. Studiene i denne litteraturgjennomgangen har brukt forskjellige former for seleksjon. Dette kan være med å påvirke denne litteraturgjennomgangens risiko for systematisk skjevhet.

Publiserings-skjevhet. Publiserings-skjevhet oppstår når studier med negative eller ikke-signifikante resultater er underrepresentert eller ikke publiseres, mens studier med positive eller signifikante resultater har større sannsynlighet for publisering. Dette fenomenet kan føre til et skjevt bilde av virkeligheten og påvirke resultatene av denne litteraturgjennomgangen. Risikoen for publiserings-skjevhet og andre typer skjevhet i inkluderte studier vurderes, men

det kan være utfordrende å fastslå tilstrekkelig informasjon for å evaluere skjevhet. I slike tilfeller betraktes resultatene og konklusjonene som usikre med hensyn til skjevhet.

Rapporteringsbias. Rapporteringsbias oppstår når studier velger å rapportere ulike utfallsmål eller utelater relevante utfallsmål i en forskningsgjennomgang. For å motvirke dette, er inklusjons- og eksklusjonskriteriene klare med hensyn til hvilke utfallsmål som er relevante. Likevel kan skjevheter oppstå hvis studiene ikke rapporterer alle relevante utfallsmål eller vurderer dem selektivt.

Resultater

Det systematiske litteratursøket med de presenterte inklusjons- og eksklusjonskriteriene førte til seks RCT-studier som ble inkludert i gjennomgangen. Fire av de seks studiene som er inkludert i denne gjennomgangen hadde som primært mål å undersøke symptomreduksjon på både angst og depresjon (Fitzpatrick et al., 2017; Fulmer et al., 2018; Klos et al., 2021; Sadeh-Sharvit et al., 2023). De to gjenværende studiene (He et al., 2022; Liu et al., 2022) hadde som primært mål å undersøke symptomreduksjon på kun depresjon. Av de seks studiene som inngår, undersøker fem av de effekten av ulike chatbots som benyttes av deltaker. Den siste studien, utarbeidet av Sadeh-Sharvit et al. (2023) undersøker effekten av en AI-plattform som benyttes av behandler som et støtteverktøy i det kliniske arbeidet. Videre sammenligner forskerne deltakernes symptomer på angst og depresjon, og ser om det er forskjeller i symptomreduksjon avhengig av om de fikk terapi med eller uten AI-plattformen.

Studiene som er inkludert i denne litteraturgjennomgangen er publisert mellom 2017 og 2023. Sentrale egenskaper ved studiene er presentert i datainnsamlingskjema i Tabell 1. Analyser og funn som ikke er relevante for problemstillingen vil ikke redegjøres for.

Tabell 1
Oversikt over sentrale egenskaper i de utvalgte studiene.

Førsteforfatter (år)	Fitzpatrick et al. (2017)	Fulmer et al. (2018)	Klos et al. (2021)	Liu et al. (2022)	He et al. (2022)	Sadeh-Sharvit et al. (2023)
Formål	Undersøker gjennomførbarhet, akseptabilitet og effekt av en fullautomatisert samtaleagent (Woebot) for å levere et selvhjelpsprogram til college-studenter som selvidentifiserte seg med symptomer på angst og depresjon.	Undersøker gjennomførbarhet og effekt av en fullautomatisert samtaleagent (Tess) for å redusere selvidentifiserte symptomer på depresjon og angst hos college-studenter.	Undersøker levedyktighet, akseptabilitet og effekt av en fullautomatisert samtaleagent (Tess), for å undersøke symptomer på depresjon og angst hos college-studenter.	Undersøker effekten av en terapichatbot (Xiaonan) som en effektiv, praktisk og skalerbar løsning for selvhjelpsterapi for depresjon hos college-studenter.	Undersøker den kliniske effekten og ikke-kliniske ytelsen til en terapichatbot (XiaoE) for unge voksne med depressive symptomer under COVID-19-pandemien hos college-studenter.	Undersøker gjennomførbarhet, akseptabilitet og effektivitet av en AI-plattform (Eleos Health) for atferdshelse med tanke på å forbedre kliniske resultater for pasienter som mottar poliklinisk behandling.
Type studie; a) antall eksperimentelle grupper, b) antall kontrollgrupper.	Åpen RCT; a) 1, b) 1.	RCT; a) 2, b) 1.	RCT; a) 1, b) 1	Åpen RCT; a)1, b)1	Forskerblindet RCT; a)1, b) 2.	RCT; a) 1, b) 1.
Deltaker-karakteristika Totalt antall, gj.sn. alder (spenn); kjønnsfordeling; etnisitet/ nasjonalitet	N=70; Gjennomsnittsalder: 22,2 år (spenn 18-28); kjønnsfordeling: 67% kvinner; etnisitet/ nasjonalitet: 79% hvite, 7% asiatiske, 9% flere etnisiteter, 2% afroamerikanske, 2% urfolk.	N=75; Gjennomsnittsalder: 22,9 år (spenn ikke oppgitt); kjønnsfordeling: 70% kvinner; etnisitet/ nasjonalitet: asiatiske: 51%, hvite: 41%, Afroamerikanske 3%, ikke-spesifisert: 5%.	N=181; Gjennomsnittsalder: ikke oppgitt (spenn 18-33); kjønnsfordeling: 87% kvinner; etnisitet/ nasjonalitet: ikke oppgitt (argentinske college-studenter).	N=83; Gjennomsnittsalder: 23,08 (spenn 19-28); kjønnsfordeling: 55,42% kvinner; etnisitet/nasjonalt: ikke spesifisert (kinesiske college-studenter).	N=148; Gjennomsnittsalder: 18.78 (17-34); kjønnsfordeling: 37.2% kvinner (55/ 148); etnisitet/ nasjonalitet: ikke spesifisert (kinesiske college-studenter).	N=47; Gjennomsnittsalder: 30,64 år (spenn ikke oppgitt); kjønnsfordeling: 72% kvinner; etnisitet/ nasjonalitet: hvite: 94% (44/47), hispanics eller latinos: 6% (3/47).
Inklusjons- og eksklusjons-kriterier	Over 18 år, evne til å lese engelsk, college- student, selvidentifiserte seg med symptomer på depresjon og angst.	Over 18 år, evne til å lese engelsk, innmeldt ved et universitet i USA, selvidentifiserte seg med symptomer på angst og depresjon.	Over 18 år, være en argentinsk innbygger, være college-student og gi informert samtykke.	18 år eller eldre, kommunisere på kinesisk, college-student, dyktige brukere av smarttelefoner, en PHQ-9-score på ni eller høyere, og ikke motta annen mental helsebehandling.	Alder 17-34 år, Gjennomsnittlig poengsum på CSMHSS mellom 2 og 3, Evne til å lese kinesisk. Eksklusjon ved høy skåre på selvmord/ vrangforestilling på CSMHSS bruk av psykiatriske medisiner.	Voksne henvist til poliklinikk med hoveddiagnose av depresjon eller angstlidelse. Eksklusjon ved alvorlig fysisk eller psykisk helseplage som kunne påvirke deltakelse/etterlevelse av behandling.

Frafall	58 av 70 rekrutterte fullførte studien, frafallsrate på 17%. Kontrollgruppen hadde en frafallsrate på 31%, Woebot hadde frafallsrate på 9%.	0% frafall fra testgruppene og mindre enn 1% frafall i kontrollgruppen (1/24).	Data ved uke 8 ble levert av 39 av de 99 (39%) deltakerne i eksperimentgruppen og 34 av de 82 (41%) i kontrollgruppen.	Samlet frafallsrate på 24,10% (20/83). 8 falt fra i intervensjonsgruppen og 12 i kontrollgruppen.	5 av 49 i XiaoE gruppen falt fra. 17 av 49 deltakere i e-bokgruppen falt fra. 20 av 50 deltakere i Xiaoai-gruppen falt fra.	Data fra 22 av 47 deltakeres selvrapporingsskjema ble analysert (11/23 i eksperimentell gruppe og 11/24 i kontrollgruppen). Oppmøtedata fra alle deltakere ble analysert.
Intervensjonskarakteristika	a) Terapichatbot (CBT), b) Plattformuavhengig chat-applikasjon, c) 12,14 sesjoner (spenn 8-18), d) 2 uker, e) uveiledet.	a) Terapichatbot (CBT, MBT), b) plattformuavhengig chat-applikasjon, c) gruppe 1: 283 meldinger (spenn 72-755), gruppe 2: 286 (spenn 133-535) d) to og fire uker, e) uveiledet.	a) Terapichatbot (CBT, EFT, MI, løsningsfokusert korttidsbehandling), b) Facebook Messenger, c) 116 (spenn 162-1290), d) 8 uker, e) uveiledet.	a) Terapichatbot (CBT), b) tekst/talemelding gjennom plattformuavhengig chat-applikasjon, c) <i>Ikke oppgitt</i> , d) 16 uker, e) uveiledet.	a) Terapichatbot (CBT), b) tekst, bilde og stemme via offisiell plattform, c) 25,54 daglige interaksjoner (spenn 0-172), d) én uke, e) uveiledet.	a) AI- plattform som hjelpemiddel for ansikt-til ansikt terapisesjoner b) Internettbasert AI-plattform, c) 5.24 økter (spenn 1-9 økter), d) 8 uker, e) ingen oppfølgingsdata.
Utfallsmål	PHQ-9; GAD-7; PANAS; Akseptabilitet, brukervennlighet (5-punkts Likert-skala og kvalitative tilbakemeldinger); totalt antall interaksjoner.	PHQ-9; GAD-7; PANAS; Bruker-tilfredshet, engasjement (skalerte og åpne spørsmål); totalt antall meldinger.	PHQ-9; GAD-7; Analyse av frafall; totalt antall meldinger; kvalitative tilbakemeldinger.	PHQ-9; GAD-7; PANAS; WAI-SR; CSQ-8; Selv-rapportert overholdelse (4 punkts Likert-skala); kvalitative tilbakemeldinger.	PHQ-9; WAQ; UMUX-LITE; Akseptabilitet (5 punkts Likert-skala); Antall og frekvens av interaksjoner; kvalitative tilbakemeldinger.	PHQ-9; GAD-7; Spørreskjema om behandlingstilfredshet og oppfattet hjelpsomhet (5 punkts Likert-skala); analyse av oppmøte; kvalitative tilbakemeldinger.
Kontrollbetingelse	Tilgang til e-bok om depresjon blant college-studenter; "National Institute of Mental Health ebook".	Tilgang til e-bok om depresjon blant college-studenter; "National Institute of Mental Health ebook".	Tilgang til e-bok om depresjon blant college-studenter; "National Institute of Mental Health ebook".	Tilgang til e-boken "Change Your Thinking - Positive and practical ways to overcome stress, negative emotions and self-defeating behavior using CBT".	1) Tilgang til e-bok om depresjon, "I Had a Black Dog". 2) chat med Xiaoai minst 1 gang om dagen (bot designet for ikke-klinisk småprat).	Mottok de vanlige tjenestene som ble tilbudt i senteret. Terapeutene kunne bruke de strategiene de mente ville være mest vellykkede.
Administrator	Woebot (uveiledet)	Tess (uveiledet)	Tess (uveiledet)	Xiaonan (uveiledet)	XiaoE (uveiledet)	Terapeuter ved Ozark Center, Missouri, USA.
Finansieringskilde;	Woebot Labs Inc. Dekket deltakerincentiver. Ellers ikke spesifisert.	X2AI Inc. dekket deltakerincentiver. Ellers ikke spesifisert.	Ingen finansiering.	Ingen finansiering.	AI4SDGs Research Program; the National Social Science Foundation of China.	Ikke oppgitt.

Karakteristika ved de psykologiske intervensjonene

Fitzpatrick et al. (2017). Deltakerne i denne studien ble tilfeldig tildelt enten en integrert psykologisk AI-chatbot ved navn Woebot eller en informasjonskontrollgruppe som fikk tilgang til å lese en e-bok om depresjon blant college-studenter. Den automatiserte samtaleagenten Woebot leverte kognitiv atferdsterapi (CBT) gjennom daglige samtaler og humørsporing. Deltakerne hadde muligheten til å bruke Woebot gjennom en plattformuavhengig chat-applikasjon på datamaskin eller mobil enhet og kunne ta kontakt med Woebot når det passet dem. Samtalene med Woebot startet med generelle spørsmål om deltakernes livssituasjon og humør, og deltakerne kunne svare med ord eller emojier for å uttrykke sin følelsesmessige tilstand. Woebot formidlet sentrale CBT-konsepter ved å inkludere lenker til korte videoer eller ordspill som hjalp deltakerne med å forstå kognitive forvrengninger. Videre hadde Woebot terapeutiske funksjoner som empatisk respons til deltakernes humør, tilpasset innhold basert på humør, etablering av personlige mål, oppfølging av tidligere aktiviteter, samt motivasjon og engasjement gjennom daglige meldinger. Woebot tilbød også refleksjon ved hjelp av ukentlige humørgrafer.

Fulmer et al. (2018). I denne studien ble deltakerne randomisert til tre forskjellige intervensjonsgrupper. Kontrollgruppen mottok informasjon i form av en e-bok om depresjon blant college-studenter. De to testgruppene mottok intervensjonen ved å samhandle med chatboten Tess, en integrert psykologisk AI-chatbot. Varigheten av intervensjonen varierte mellom gruppene, der testgruppe 1 samhandlet med Tess i 2 uker, mens testgruppe 2 hadde en lengre intervensjonsperiode på 4 uker. Tess formidlet innholdet i intervensjonen gjennom individuelt tilpassede samtaler, hvor ulike psykologiske tilnærminger, inkludert kognitiv atferdsterapi og mindfulness-basert terapi ble anvendt. Chatboten skulle gi deltakerne en omfattende intervensjon som inkluderte psykologisk støtte, psykoedukasjon og målrettede påminnelser. Deltakerne i testgruppene kunne engasjere seg i samtaler med Tess når som

helst i løpet av intervensjonsperioden. Tess tilpasset seg deltakernes individuelle behov og emosjonelle tilstand ved hjelp av avanserte naturlig språkforståelsesalgoritmer. Chatboten leverte informasjon om depresjon, angst og mestringsstrategier, og oppfordret deltakerne til å utforske og reflektere over egne følelser og erfaringer. Tess ga også praktiske råd angående søvn, kosthold og fysisk aktivitet for å forbedre deltakernes generelle velvære. Gjennom dialogene ble det lagt vekt på å bygge en emosjonell forbindelse med deltakerne, og chatboten brukte forskjellige motiverende teknikker for å opprettholde engasjementet.

Klos et al. (2021). Deltakerne i denne studien ble tilfeldig tildelt enten eksperimentell gruppe, som hadde tilgang til den integrerte AI-baserte chatboten ved navn Tess, eller en kontrollgruppe, som mottok en elektronisk bok om depresjon blant college-studenter. Chatboten tilbød støtte for mental helse gjennom tekstbaserte samtaler. Tess kommuniserte med deltakerne via Facebook Messenger og tilbød daglige eller annenhverdags samtaler over en 8-ukers intervensjonsperiode. Samtalene med Tess ble skreddersydd etter deltakernes individuelle behov og bygget på flere teoretiske tilnærminger, inkludert kognitiv atferdsterapi, emosjonsfokusert terapi, løsningsfokusert korttidsbehandling, og motiverende samtale. Tess brukte forhåndsdefinerte uttalelser for å reagere på deltakernes uttrykte følelser og bekymringer, og tilbød støttende respons som var tilpasset deltakernes behov. For eksempel, dersom en deltaker uttrykte angst, ville Tess respondere med en avslapningsstrategi. Tess benyttet seg av emoji'er for å sikre en brukervennlig opplevelse.

Liu et al. (2022). Formålet i denne studien var å sammenligne effekten av chatbot-terapi med en kontrollgruppe som mottok en CBT-basert selvhjelpsbok. Chatbot-intervensjonen ble presentert gjennom en terapichatbot ved navn XiaoNan, og hadde en varighet på 16 uker. XiaoNan var forankret i prinsippene for kognitiv atferdsterapi og var trent for å lindre depresjon ved å hjelpe brukere med å forstå og transformere tankemønstre. Chatboten bistod blant annet brukere med å avdekke automatiske negative tanker, utforske

sammenhengen mellom tanker, følelser, og atferd, og med å etablere nye og mer adaptive tankemønstre. Videre tilbød chatboten daglige følelsesoppdateringer og spontane samtaler, slik at deltakerne jevnlig kunne sjekke inn for å reflektere over følelsesmessige endringer og få støtte etter behov. Chatboten var også i stand til å gi empatisk støtte og oppfordre deltakerne til å søke hjelp fra profesjonelle terapeuter hvis det ble ansett som nødvendig. Chatbot-terapien ble levert gjennom en interaktiv digital plattform, tilgjengelig via smarttelefoner eller datamaskiner. Deltakerne i denne gruppen kunne kommunisere med XiaoNan ved å sende tekst- eller talemeldinger gjennom plattformen.

He et al. (2022). Deltakerne i denne studien ble randomisert til tre ulike grupper. Den eksperimentelle gruppen fikk tilgang til chatboten XiaoE. Den første kontrollgruppen hadde tilgang til en e-bok om depresjon, mens den andre kontrollgruppen hadde tilgang til en generell chatbot kalt Xiaoai. Intervensjonen strakte seg over én uke. XiaoE, en chatbot basert på kognitiv atferdsterapi, ble spesielt utviklet for å adressere depresjon ved å tilby screening, forebygging og selvhjelp gjennom tekst-, bilde- og stemmebasert interaksjon med brukerne. Den var forankret i CBT-prinsippene og besto av syv moduler som dekket ulike psykologiske konsepter. Deltakerne i XiaoE-gruppen ble oppfordret til å gjennomføre én modul daglig i løpet av intervensjonsperioden, med en ekstra modul kalt "Takk" for å fremme positive hendelser og humør. XiaoE inkluderte også prosessveiledning og daglige påminnelser om oppgaver. Under studien ble deltakerne kun veiledet av XiaoE, mens personalet kun besvarte tekniske eller operative spørsmål. I e-bokgruppen ble deltakerne bedt om å lese en e-bok om depresjon, og de mottok daglige artikler om depresjon som korresponderte med modulene i XiaoE-gruppen. I Xiaoai-gruppen ble deltakerne oppmuntret til å kommunisere med Xiaoai minst én gang daglig. Samtalen mellom deltakerne og Xiaoai var ubegrenset i innhold, men daglige samtaleemner ble foreslått i tråd med temaene i intervensjonsgruppen.

Sadeh-Sharvit et al. (2023). Denne studien inkluderte 47 voksne som mottok individuell kognitiv atferdsterapi ved en samfunnsbasert klinikk i USA. Intervensjonen ble administrert av terapeuter, og deltakerne ble tilfeldig tildelt enten terapi støttet av Eleos Healths AI-plattform eller standardbehandling (Treatment as Usual - TAU) innenfor samme klinikk. Terapeutene hadde tilgang til Eleos Health-plattformen i de første 2 månedene av behandlingen. Plattformen hadde flere funksjoner, inkludert overvåking og transkripsjon av terapisesjoner, tilbakemelding til terapeuter om bruk av evidensbaserte metoder, og integrasjon av disse dataene med rutinemessige spørreskjemaer utfylt av pasientene. Innholdet i intervensjonene ble levert som en støttende verktøy i tillegg til tradisjonell terapi. Innholdet baserte seg på psykologisk teori og evidensbaserte metoder for behandling av depresjon og angstlidelser, og terapeutene kunne bruke informasjonen som AI-plattformen ga på den måten de selv ønsket.

Kontrollbetingelser

Studiene som er inkludert i denne litteraturgjennomgangen benytter seg av ulike kontrollbetingelser. I Fitzpatrick et al. (2017), Klos et al. (2021) og Fulmer et al. (2018) sine studier ble informasjonskontrollgruppen henvist til en e-bok om depresjon blant college-studenter, "National Institute of Mental Health eBook". I Liu et al. (2022) sin studie ble også kontrollgruppen bedt om å lese en selvhjelpsbok om kognitiv terapi, kalt "Change Your Thinking - Positive and practical ways to overcome stress, negative emotions and self-defeating behavior using CBT". I He et al. (2022) sin studie var det én kontrollgruppe som mottok en e-bok om depresjon, "I Had a Black Dog" og en annen kontrollgruppe som fikk tilgang til en generell chatbot som ikke var rettet mot psykisk helse, men som ga dagligdagse, ikke-kliniske samtaler. I Sadeh-Sarvit (2023) sin studie ble en eksperimentell gruppe som mottok terapi med støtte fra en AI-plattform sammenlignet med en kontrollgruppe som mottok "treatment as usual" ved en poliklinikk i USA.

Utfallsmål

The Patient Health Questionnaire-9 (PHQ-9) ble brukt som evalueringsverktøy i alle studiene. PHQ-9 er et selvrappoterings skjema med ni elementer som måler hyppigheten og alvorligheten av depressive symptomer de siste to ukene. Dette verktøyet er blant de mest anvendte, reliable og validerte kartleggingsverktøyene for depressive symptomer. Hvert av de ni elementene bygger på kriteriene for depressiv lidelse i Diagnostic and Statistical Manual of Mental Disorders, 4. utgave (DSM-IV). Skalaen spenner fra null (ingen symptomer) til tre (nesten hver dag), og poengsummene varierer fra 0-5 (ingen symptomer) til 5-9, 10-14, 15-20 og 20, som representerer henholdsvis ingen symptomer, mild, moderat, moderat-alvorlig og alvorlig depresjon.

Generalisert angstlidelse-7 (GAD-7) er benyttet som utfallsmål i fem av studiene (Fitzpatrick et al., 2017; Fulmer et al., 2018; Klos et al., 2021; Liu et al., 2022; Sadeh-Sharvit et al., 2023). GAD-7 er et validert, kort selvrappoteringsverktøy for å vurdere hyppigheten og alvorlighetsgraden av engstelige tanker og atferd i løpet av de siste to ukene. Basert på DSM-IV-diagnostiske kriterier for GAD, spenner poengene for alle syv elementene fra 0 (slett ikke) til 3 (nesten hver dag). En poengsum på 10 indikerer moderat angst, mens en poengsum større enn 15 indikerer alvorlig angst.

PANAS - Positive and Negative Affect Schedule er benyttet i tre av studiene (Fitzpatrick et al., 2017; Fulmer et al., 2018; Liu et al., 2022). Skalaen for positiv og negativ affekt er en selvrappoteringsmåling med 20 elementer som måler nåværende positiv og negativ affekt. Halvparten av elementene representerer positiv affekt, mens den andre halvparten av elementene indikerer negativ affekt. Elementene skåres på en skala fra 1 (svært lite eller overhodet ikke) til 5 (ekstremt), med høyere poengsummer som representerer høyere affekt. Positiv og negativ affekt summeres uavhengig av hverandre med mulige poengsummer fra 10-50.

Samtlige studier i denne litteraturgjennomgangen har som primært mål å undersøke effekt av intervensjonene gjennom validerte selvrappoteringskjema. I tillegg til standardiserte, validerte selvrappoteringsmål benytter studiene seg av andre utfallsmål, presentert i Tabell 1. Sekundære mål i studiene varierer fra å undersøke gjennomførbarhet og akseptabilitet (Fitzpatrick et al., 2017; Fulmer et al., 2018; Klos et al., 2021; Sadeh-Sharvit et al., 2023) til klinisk og ikke-klinisk ytelse og terapeutisk allianse, også kalt arbeidsallianse (He et al., 2022; Liu et al., 2022). Samtlige studier i denne litteraturgjennomgangen har samlet inn kvalitative tilbakemeldinger, enten via åpne kommentarer eller fem punkts Likert-skala, og analysert oppmøte eller interaksjoner. Studiene har også samlet data for å vurdere brukervennlighet, brukertilfredshet, engasjement, selvrappotert overholdelse og oppfattet hjelpsomhet.

Effekt på symptomreduksjon

Tre av studiene som er inkludert i denne gjennomgangen undersøkte chatbots sin effekt på symptomer på både angst og depresjon. I studien til Fitzpatrick et al. (2017) ble det funnet en moderat reduksjon av depresjonssymptomer blant deltakerne, med en Cohen's d-effektstørrelse på 0.44. Det ble ikke observert signifikante endringer i angstsymptomer. For å kartlegge kilden og omfanget av endringer i depresjon, ble gjentatte målinger for enkeltspørsmål i PHQ-9 gjennomført. Endringer ble observert på følgende spørsmål i synkende rekkefølge av omfang: motoriske symptomer ($d=2.09$), appetitt ($d=0.65$), liten interesse eller glede over ting ($d=0.44$), dårlig følelse om seg selv ($d=0.40$), konsentrasjon ($d=0.39$), selvmordstanker ($d=0.30$), følelse av nedstemthet ($d=0.14$), søvn ($d=0.12$) og energi ($d=0.06$). Resultatene i Fulmer et al. (2018) sin studie viste en signifikant reduksjon i både depressive symptomer og angstsymptomer hos de eksperimentelle gruppene, sammenlignet med kontrollgruppen. Disse effektene er ikke tallfestet i studien. Det ble observert en moderat til stor effektstørrelse for depresjon. Effekten på angstsymptomer blir beskrevet som

signifikant, men er ikke ytterligere spesifisert. I studien til Klos et al. (2021) observert en signifikant nedgang i angstsymptomer i eksperimentgruppen ($d = 0.50$) fra start til uke 8, mens kontrollgruppen ikke viste endringer i angstnivå. Ingen signifikante endringer ble funnet for depressive symptomer, hverken i eksperiment- eller kontrollgruppen. Den eksperimentelle gruppen opplevde en betydelig reduksjon i angstsymptomer og en nær-signifikant trend for depressive symptomer etter 8 uker med intervensjon. Resultatene avviker fra tidligere studier (Fitzpatrick et al., 2017; Fulmer et al., 2018), ved at det ikke ble funnet signifikant effekt for depressive symptomer.

To av studiene hadde som primært mål å undersøke effekt på kun depresjon. He et al. (2022) beskriver at depressive symptomer i XiaoE-gruppen ved T1, umiddelbart etter én ukes behandling, hadde en signifikant reduksjon sammenlignet med kontrollgruppene, med en stor effektstørrelse ($d=0.51$). Den langvarige effekten ved T2, 1 måned etter intervensjonen, var fortsatt betydelig, selv om effektstørrelsen var redusert til medium størrelse ($d=0.31$). I studien av Liu et al. (2022) opplevde deltakerne i intervensjonsgruppen en betydelig reduksjon i symptomer på depresjon. Selv om det ikke var et primært forskningsspørsmål, ble mål på angst inkludert, og effekten på angst var moderat. Effekten på reduksjon av depresjon var høy ($d = 0.83$), mens effekten på reduksjon av angst var moderat ($d= 0.30$).

Resultatene i studien til Sadeh-Sharvit et al. (2023) demonstrerer positive effekter av Eleos Health's AI-plattform som støtteverktøy i ansikt-til-ansikt terapi, sammenlignet med standardbehandling (TAU) for pasienter som mottar poliklinisk behandling for depresjon eller angstlidelse. Pasientene som mottok AI-støttet terapi opplevde en signifikant økning i oppmøtefrekvensen (67%) sammenlignet med TAU-gruppen. Videre ble det observert en betydelig reduksjon i depresjonssymptomer (34% reduksjon) og angstsymptomer (29% reduksjon) i AI-gruppen, med store effektstørrelser ($d = 0.82$ for depresjon, $d = 0.78$ for angst).

Tabell 2

Signifikante funn på post-intervensjonsmålinger. Effektstørrelse er oppgitt i de tilfellene det er spesifisert i studien.

Studie	Depresjon (PHQ-9)	Angst (GAD-7)
Fitzpatrick et al. (2017)	Ja (d = 0.44)	Nei (ikke signifikant)
Fulmer et al. (2018)	Ja (moderat-stor)	Ja (uspesifisert)
Klos et al. (2021)	Nei	Ja (d = 0.50)
Liu et al. (2022)	Ja (d = 0.83)	Ja (d = 0.30)
He et al. (2022)	Ja (d = 0.51)	Ikke undersøkt.
<i>Effekt etter 1 mnd: d = 0.31.</i>		
Sadeh-Sharvit et al. (2023)	Ja (d = 0.82)	Ja (d = 0.78)

Tilbakemeldinger, akseptabilitet og engasjement

Fitzpatrick et al. (2017). Tilbakemeldinger, akseptabilitet og engasjement ble vurdert av begge betingelser ved å bruke spørsmål med blandet format. Deltakerne vurderte total tilfredshet og innholdstilfredshet på en 5-punkts Likert-skala, samt graden av fremmet emosjonell bevissthet, læringens relevans for dagliglivet, og ga tilbakemelding på beste og verste opplevelse. Hovedfokuset var på kvalitative svar for Woebot-betingelsen, men spørsmål til kontrollgruppen ga også innsikt i engasjement. Woebot-deltakere viste betydelig høyere tilfredshet, økt følelsesmessig bevissthet, og 100% bekreftet læring, sammenlignet med 77% i kontrollgruppen. Ingen forskjeller ble funnet i oppfatningen av læringens relevans for hverdagen. Woebot-deltakerne engasjerte seg i gjennomsnitt 12 ganger i løpet av intervensjonsperioden. Kvalitative tilbakemeldinger inkluderte positiv respons på Woebot, med nevning av ansvarsfølelse, empati, verdifull læring og personifisering av Woebot. Deltakerne beskrev blant annet Woebot med ord som "han", "en venn" og "en morsom, liten fyr", som de hadde utviklet en relasjon til. Deltakerne i Woebot-gruppen uttrykte høy tilfredshet, økt emosjonell bevissthet og positivt læringsutbytte.

Fulmer et al. (2018). Engasjement i studien ble målt gjennom antall meldinger utvekslet mellom Tess og deltakerne. Engasjement ble vurdert som høyt. Gruppe 1 og gruppe 2 utvekslet i gjennomsnitt henholdsvis 283 og 286 meldinger, og deltakere indikerte høy tilfredshet og relevans for deres daglige liv. I tillegg ga deltakerne kvalitative tilbakemeldinger, hvor deltakerne ble spurt om prosess og innhold, og hva som var det beste og verste med Tess. Prosesstemaer inkluderte ansvarlighet, empati og læring, mens problemene var knyttet til begrensninger i naturlige samtaler og manglende interaktivitet. Totalt mottok studien 48 åpne kommentarer fra 50 deltakere, og engasjement og tilfredshet var høyere blant de som samhandlet med Tess sammenlignet med kontrollgruppen som brukte e-boken. Deltakerne viste høy grad av tilfredshet og engasjement med Tess, og rapporterte at Tess ga mer personlig støtte enn tradisjonelle terapimetoder.

Klos et al. (2021). I denne studien ble engasjement i hovedsak vurdert ut ifra antall meldinger som ble utvekslet med chatboten. Gjennomsnittlig antall utvekslede med tess var 472, med et spenn fra 162 til 1290 meldinger, etter 8 uker. Tilbakemeldinger fra deltakerne ble også samlet inn etter hver samtale ved å spørre deltakeren om samtalen var nyttig. Svarene ble kodet som positive, negative eller ambivalente. Resultatene viste at et høyere antall meldinger utvekslet med Tess var assosiert med positiv tilbakemelding, og forskerne tolket dette som en sammenheng mellom engasjement og positiv opplevelse blant deltakerne. Majoriteten av tilbakemeldingene på om samtalen var nyttig ble rangert som positive.

Liu et al. (2022). Denne studien undersøkte overholdelsesrater, brukertilfredshet og terapeutisk allianse for å kunne si noe om hvordan intervensjonene virket. Klienttilfredshet og selvrapporert overholdelse var lik i eksperimentell og kontrollgruppe. Imidlertid ble det funnet en nedadgående trend i overholdelsesraten for chatboten, noe som kan ha skyldtes tekniske og innholdsmessige feil ved chatboten. WAI-SR-scorene til chatbot-brukerne var høyere, som altså er mål på terapeutisk allianse. Økt selvavsløring ble også observert i

tilbakemelding fra brukerne der uttrykk for økt selvavsløring ble funnet. I tillegg viste tilbakemeldingen om XiaoNan at prosessfaktorer var viktigere enn innhold.

He et al. (2022). Deltakerne som brukte XiaoE opplevde bedre samarbeidsmestring og akseptabilitet sammenlignet med kontrollgruppene. Bedre akseptabilitet ble observert gjennom total skåre på Acceptability Scale (AS), tilfredshet med innhold, følelsesmessig bevissthet, læring av ny kunnskap og relevans til dagliglivet. Intervensjonsgruppen viste også høyere poengsum på Working Alliance Questionnaire (WAQ) totalt, spesielt på subskalaene "Bond" og "Engagement". Brukervennligheten var omtrent den samme i alle gruppene. XiaoE-gruppen utviste høyt engasjement gjennom hele studien, og hadde høyere daglig interaksjonsfrekvens og varighet enn i kontrollgruppene. Temaene forhold, følelser, tilpasning og praktisk bruk ble identifisert som positive aspekter basert på deltakernes tilbakemeldinger, mens temaene "innhold" og "teknologi" ble nevnt som negative opplevelser. Ønsker om forbedringer gjaldt en mer flytende dialogprosess, økt følelsesmessig respons og samhandling, samt serveroppgraderinger. Antropomorfering av chatboten ble også funnet i tilbakemeldinger, deriblant ble det sagt at XiaoE var "XiaoE er veldig søt, jeg liker å snakke med XiaoE, han vil være med og akseptere meg, så jeg føler meg ikke ensom". Sammenlignet med tidligere arbeid understreker denne studien betydningen av å undersøke spesialiserte chatbots som XiaoE for å lindre depressive symptomer, i motsetning til generelle chatbots som Xiaoai.

Sadeh-Sarvit et al. (2023). Pasienter i begge grupper rapporterte høy tilfredshet med terapien og opplevde den som svært hjelpsom for å oppnå sine mål, og ingen signifikante gruppeforskjeller ble funnet etter to måneders behandling. Pasienter som mottok terapi med AI-plattformens støtte hadde bedre oppmøte til terapitimer, større reduksjon i depresjon og angstsymptomer, og terapeutene brukte mindre tid på dokumentasjon sammenlignet med TAU-gruppen. Spesifikt hadde pasientene i AI-gruppen en 67% høyere oppmøtefrekvens

sammenlignet med TAU-gruppen. Det ble ikke funnet noen gruppeforskjell når det gjelder tilfredshet med behandlingen og oppfattet hjelpsomhet etter to måneders behandling. Videre la terapeutene som brukte AI-plattformen, inn sine fremdriftsnotater i gjennomsnitt 55 timer tidligere enn terapeutene i TAU-gruppen, og rapporterte om at de ble mer bevisst på sine egne CBT-ferdigheter.

Frafallsrater

Studiene i denne litteraturgjennomgangen rapporterer om stor variasjon i sine frafallsrater. Frafallsratene ble i samtlige studier beregnet ut ifra gjenstående deltakerantall etter manglende overholdelse av utfylling av validerte selvrapporeringsmål etter endt intervensjon. Det er store variasjoner i frafall mellom studiene, men også internt i de forskjellige gruppene i studiene.

I Fitzpatrick et al. (2017) sin studie var det en total frafallsrate på 17%. Det var variasjon mellom de forskjellige gruppene når det gjaldt frafall. Kontrollgruppen hadde en frafallsrate på 31%, og den eksperimentelle gruppen hadde frafallsrate på 9%. I studien til Fulmer et al. (2018) var det 0% frafall fra testgruppene og mindre enn 1% frafall i kontrollgruppen. I studien utført av Klos et al. (2021) var frafallet blant deltakerne 61% i eksperimentgruppen og 59% i kontrollgruppen. I Liu et al. (2022) sin studie var det en samlet frafallsrate på 24,10%, hvor 8/41 deltakere falt fra i testgruppen og 12/40 i kontrollgruppen. I He et al. (2022) sin studie var det et frafall på 10% i testgruppen, 35% i e-bokgruppen og 40% i den generelle chatbotgruppen. I studien til Sadeh-Sharvit et al. (2023) ble selvrapporeringsskjemaet til 53% av deltakerne analysert (11/23 i eksperimentell gruppe og 11/24 i kontrollgruppen). Deltakere som ikke fylte ut selvrapporeringsskjemaene fikk fortsette å delta i studien, og tolkninger på andre aspekter enn symptomreduksjon ble gjennomført på et høyere antall deltakere.

Diskusjon

Oppsummering av funn

Det primære forskningsspørsmålet i denne litteraturgjennomgangen var hvorvidt psykologisk terapi basert på kunstig intelligens virker mot symptomer på angst og depresjon. Samlet sett indikerer resultatene at intervensjoner basert på kunstig intelligens har potensial til å effektivt redusere depressive symptomer ($d = 0.44 - 0.83$), samtidig som effektene på angst kan variere ($d = 0.30 - 0.78$). Denne gjennomgangen av RCT-studier antyder at chatbots kan være nyttige i behandlingen av depresjon, men ytterligere forskning er nødvendig for å forstå og forbedre deres effektivitet, særlig for angstlidelser.

Det sekundære spørsmålet av interesse i denne litteraturgjennomgangen var hvordan behandling basert på kunstig intelligens har virkning, og hvorvidt det er en sammenheng med teori på psykologisk avstand og antropomorfisme. Resultatene fra denne gjennomgangen, inkludert tilbakemeldinger, engasjement, akseptabilitet, frafallsrater og karakteristika ved design og innhold, tyder på at chatbots kan skape opplevelser som i visse tilfeller ligner menneskelig interaksjon, spesielt når det gjelder å håndtere depresjon. Disse funnene må samtidig tolkes med forsiktighet, særlig fordi akseptabilitet, engasjement, tilbakemeldinger og frafallsrater varierte betydelig mellom studiene.

Utfallsmål på symptomreduksjon

Når det gjelder virkningen av AI-terapi mot symptomer på angst og depresjon, er det særlig én studie sine som skiller seg ut i denne litteraturgjennomgangen. Klos et al. (2021) fant ingen signifikant effekt på depresjon ved bruk av chatboten Tess. Dette skilte seg fra tidligere studier, som inkluderer Fitzpatrick et al. (2017) og Fulmer et al. (2018), som observerte en betydelig effekt på depresjon. Det er viktig å merke seg at tidligere RCT-studier har oppgitt sin effekt på forskjellige måter, der de enten ikke har tallfestet effekten av sine funn eller tolket de på en annen måte, som gjør sammenligning utfordrende. Klos et al. (2021)

gir imidlertid flere forklaringer på avvikene i deres resultater i forhold til tidligere forskning:

1) Studien var underdrevet, 2) selv om funnene ikke var statistisk signifikante, indikerte de en lignende tendens som tidligere studier, 3) det kan være nødvendig med høyere doser av lavintensitetsinngrep fra chatbots for å oppnå tydelige mellomgrupppeeffekter blant argentinske studenter, 4) variasjon i Tess sine samtaler basert på ulike teoretiske tilnærminger kan ha redusert den terapeutiske kraften, og 5) kvaliteten på intervensjonen kan ha blitt redusert under oversettelse fra engelsk til spansk.

Studiene som er med i denne litteraturgjennomgangen har som mål å finne ut om chatbots og AI-plattformer kan redusere symptomer på angst og/eller depresjon. Samtidig, er det kun én av studiene som går i dybden på *hvilke enkeltsymptomer* på angst og depresjon som er redusert (Fitzpatrick et al., 2017), basert på de kartleggingsinstrumentene som er besvart av deltakerne. Majoriteten av studiene presenterer kun økning eller nedgang som gjennomsnittsmålinger ut ifra PHQ-9 og GAD-7. Det er derfor vanskelig å stadfeste om det er noen forskjell i hvilke spesifikke symptomer som har blitt redusert.

Det er også forskjeller mellom studiene når det gjelder hvordan spørreskjemaene har blitt tolket, og dette kan utgjøre risiko for systematisk skjevhet eller feiltolkning av resultatene. Klos et al. (2021) har målt symptomer på depresjon ved hjelp av PHQ-9 som en kategorisk og ordinal variabel, mens Fulmer et al. (2018) og Fitzpatrick et al. (2017) brukte det som en kontinuerlig måling. Fulmer et al. (2018) tallfestet heller ikke sine funn. Dette begrenser muligheten for å sammenligne effekten av resultatene, og kan øke risikoen for systematisk skjevhet i rapporteringen av resultatene. Flere av studiene nevner ikke eksplisitt hvordan resultatene er avlest og tolket, som også gjør det utfordrende å vurdere om det kan være publiseringskjevhet i studiene. Studiene sammenligner dessuten mål basert på selvrapportering, som også kan bidra til å påvirke resultatene i studien, særlig siden ingen av deltakerne var blindet.

Antropomorfisme og psykologisk avstand

Det er et sekundært spørsmål av interesse i denne litteraturgjennomgangen å finne ut av hvilke underliggende mekanismer som kan stå bak virkningen av AI for å redusere symptomer på angst og depresjon. Psykologisk avstand og antropomorfisme har blitt foreslått som medierende faktorer bak denne virkningen. Chatbots forsøker ofte å speile en menneskelig terapeutisk prosess, samtidig som de leverer oppgaveorienterte og evidensbaserte intervensjoner. Det finnes en del menneskelige faktorer ved chatbots, blant annet ved at de svarer i sanntid, kan inneholde ansikt, stemme, og at de benytter empatiske responser (Ma et al., 2019). Design av chatbots blir også mer og mer menneskelignende, der de vanligste indikatorene på menneskelighet for chatbot-agenter inkluderer visuelle, identitets- og konversasjonelle indikatorer (Go & Sundar, 2019). Det er ikke gjort eksplisitte mål på psykologisk avstand og antropomorfisme i studiene inkludert i denne gjennomgangen, men kvalitative tilbakemeldinger fra deltakere, samt mål på arbeidsallianse og designmessige antropomorfe egenskaper, kan bidra til å forstå funnene basert på teori om psykologisk distanse og antropomorfisme.

Tilbakemeldinger fra deltakere i flere studier kan tyde på at antropomorfisme har redusert den psykologiske avstanden i interaksjonen. Flere av de kvalitative tilbakemeldingene i studiene viser at deltakerne personifiserte chatboten de snakket med. I studien til He et al. (2022) ble det følgende tilbakemeldt fra en deltaker: "XiaoE er veldig søt, jeg liker å prate med XiaoE, han vil være med og akseptere meg, så jeg føler meg ikke ensom.". Fitzpatrick og kolleger (2017) mottok tilbakemeldinger fra sin testgruppe der Woebot ble referert til med personkarakteristika som "han", "en venn" og en "morsom liten fyr". Forskerne formidler samtidig at de ga chatboten navnet Woebot for å understreke samtaleagentens ikke-menneskelige karakter, og oppgir at det derfor er bemerkelsesverdig at oppfattet kilde til empati av deltakerne var selve chatboten, og ikke utviklerne som stod bak.

Det finnes også designmessige aspekter ved enkelte av studiene som kan ha bidratt til å øke grad av antropomorfisme og redusere psykologisk avstand. For eksempel bruker Chatboten Tess emojis for å komme nærmere sine brukere, som er vist i studien til Chung og Han (2022) å redusere psykologisk avstand. Chatbotene som er benyttet i disse studiene tillater tilpasning av innhold, språk og kommunikasjonskanaler for spesifikke behandlingsformer eller brukergrupper. Denne tilpasningen bidrar sannsynligvis til en mer personlig og brukersentrert opplevelse. De bruker også emosjonsalgoritmer og maskinlæringsteknikker for å støtte ulike funksjoner. Integreringen av emosjonsidentifikasjon antyder et forsøk på å etterligne menneskelig emosjonell forståelse, noe som kan bidra til antropomorfisme og redusere psykologisk distanse. Flere av Chatbotene kunne også konfigureres til å levere tjenester via eksisterende kommunikasjonskanaler, som for eksempel Facebook Messenger. Denne tilnærmingen øker tilgjengeligheten og integrerer chatboten i brukernes eksisterende kommunikasjonsvaner, og fremmer potensielt en mer naturlig interaksjon. Alle chatbotene var i stand til å tolke frie tekstmeldinger, slik at brukere kunne kommunisere med en naturlig flyt i samtalen. Denne funksjonen reflekterer et forsøk på å etterligne menneskelig kommunikasjon.

Samlet sett tyder resultatene fra denne litteraturgjennomgangen på at chatbots kan skape opplevelser som i visse tilfeller ligner menneskelig interaksjon, spesielt når det gjelder å håndtere depresjon. En mulig måte å forstå dette på er ved å tilskrive antropomorfisme og aspekter ved psykologisk distanse som mulig årsak til dette. Imidlertid må det også bemerkes at effektene var mer varierte når det gjelder angstsymptomer, noe som kan indikere at psykologisk distanse fremdeles er til stede i slike situasjoner. Ytterligere forskning er nødvendig for å forstå i hvilken grad chatbot-interaksjoner ligner menneskelig interaksjon og hvilke underliggende mekanismer som forklarer virkningen av disse aspektene. Ut ifra undertegnede (LS) sin kunnskap, finnes det til nå ingen studier som sammenligner chatbots

med menneskelige egenskaper med faktiske mennesker, som gjør det utfordrende å stadfeste hvorvidt kontakt med chatbots er like effektivt som kontakt med virkelige mennesker. Det oppfordres til at fremtidig forskning undersøker dette nærmere.

Frafallsrater

Det er store variasjoner i frafallsratene i de utvalgte studiene i denne litteraturgjennomgangen. Frafallsratene varierer fra mindre enn 1% frafall (Fulmer et al., 2018) til over 50% (Sadeh-Sharvit et al., 2023). Samlet oversikt over frafall er presentert i tabell 1. En sentral observasjon når det gjelder frafallsrater er at fire av seks studier hadde bemerkelsesverdig lav frafallsrate i intervensjonsgruppene, sammenlignet med kontrollgruppene (Fitzpatrick et al., 2017; Fulmer et al., 2018; He et al., 2022; Liu et al., 2022). Flere av studiene ser dette som en antydning til at brukerne engasjerte seg i de selvstyrte AI-intervensjonene.

Internettbaserte psykologiske intervensjoner har blitt kritisert for lav bruk og høyt frafall, og det er sett i sammenheng med mangel på veiledning og tilbakemelding. Mangelen på kontakt med en terapeut og rene selvstyrte programmer har vist seg å redusere overholdelsesratene i studier der intervensjoner gjennomføres uten aktiv veiledning eller tilsyn, også kalt uveiledede intervensjoner (Borghouts et al., 2021; Donkin et al., 2011). Forskning antyder at veiledning eller ulike former for støtte øker engasjementet i studier, og frafallsrater er høyest for intervensjoner som er utelukkende selvstyrte (Richards & Richardson, 2012). Veiledede intervensjoner, som refererer til terapeutiske eller behandlingsmessige tiltak som gjennomføres med aktiv veiledning, tilsyn eller ulike typer støtte, har en tendens til å vise større forbedringer i resultatene for mental helse i internettbaserte digitale intervensjoner. Samtidig kan veiledning redusere skalerbarheten til disse intervensjonene gitt mangelen på fagfolk innen mental helse (Boucher et al., 2021). Det er derfor hensiktsmessig å undersøke i hvilken grad chatbots kan etterligne effekten av

veiledede intervensjoner, på tross av at de som regel er uveiledet og fullstendig selvstyrte, ofte med unntak av dersom man trenger teknisk hjelp.

Terapichatbots har vist seg å ha en betydelig høyere overholdelsesrate (Vaidyam et al., 2019) enn tidligere internettbaserte psykologiske intervensjoner. Et perspektiv på hvorfor terapichatbots viser seg å ha en høy overholdelsesrate, på tross av at de er uveiledet, kan være på grunn av de antropomorfe egenskapene til en chatbot, som gjør at de inneholder menneskelige egenskaper og terapeutiske prosessfaktorer på tross av at de ikke er veiledet. Funnene fra denne litteraturgjennomgangen viste at chatbot-intervensjoner som involverer Woebot, Tess, XiaoNan, og XiaoE, har resultert i betydelige reduksjoner i depressive symptomer, og overholdelsesratene i chatbot-intervensjonene var generelt høye, særlig sammenlignet med kontrollgruppene. Dette kan tyde på at deltakerne kanskje har opplevd chatbotene som mer enn bare verktøy eller maskiner, og i stedet har utviklet en form for relasjon med dem. Denne relasjonen kan være en refleksjon av antropomorfisme, der deltakerne ser chatbotene som forståelsesfulle og emosjonelt støttende enheter, som igjen kan redusere den psykologiske distansen.

Terapeutisk allianse

Terapeutisk allianse, eller arbeidsallianse, representerer det samarbeidsmessige og emosjonelle forholdet mellom klient og terapeut, og regnes som en sentral fellesfaktor i utfallene av psykoterapi (Cuijpers et al., 2019). Terapeutisk allianse regnes også som en viktig faktor for å vurdere forholdet mellom en klient og en teknologisk samtaleagent (Darcy et al., 2021). Hvorvidt det er mulig å etablere en terapeutisk allianse med en samtaleagent på kunstig intelligens, er adressert av flere studier i denne litteraturgjennomgangen, både gjennom standardiserte mål på arbeidsallianse (He et al., 2022; Liu et al., 2022) og gjennom tolkninger av tilbakemeldinger og frafallsrater.

Terapeutisk allianse, antropomorfisme og psykologisk avstand. Ut ifra tilbakemeldinger Fitzpatrick et al. (2017) mottok i sin studie hvor deltakerne personifiserte studiens chatbot, ble det formidlet en antydning om at digitale samtaleagenter basert på kunstig intelligens kan tilnærme seg terapeutiske prosessfaktorer, og forskerne oppfordret til at dette undersøkes i fremtidige studier med standardiserte mål på arbeidsallianse. Studier som inkluderer forsøk med psykologiske chatbots har senere benyttet standardiserte mål på arbeidsallianse som utfallsmål i sine studier (Hauser-Ulrich et al., 2020; He et al., 2022; Liu et al., 2022; Prochaska et al., 2021). Flere av studiene i denne litteraturgjennomgangen påpeker at de fant at prosessfaktorer var mer innflytelsesrike enn selve innholdet, basert på kvalitative tilbakemeldinger (Fitzpatrick et al., 2017; He et al., 2022; Liu et al., 2022). I Liu et al. (2022) sin studie ble det også gjennom tilbakemelding observert uttrykk for høyere selvavsløring hos testgruppen, uten at dette ble ytterligere utdypet. Denne økte selvavsløringen, kombinert med viktigheten av prosessfaktorer, kan potensielt være et uttrykk for innflytelse av antropomorfisme og redusert psykologisk avstand.

Akkurat som at egenskapene og prosessfaktorene til chatbotene antas å formidle mye av variasjonen i positive resultater på tvers av terapimetoder, poengterer Fitzpatrick et al. (2017) at prosessfaktorer knyttet til samtalebaserte agenter, som evnen til å formidle empati, kan være i stand til både å forsterke og motsatt, bryte en terapeutisk prosess (Fitzpatrick et al., 2017). Dette er i tråd med annet arbeid som antyder at en terapeutisk allianse kan etableres mellom mennesker og ikke-menneskelige agenter i løpet av kort tid (Bickmore et al., 2005). En studie viste dessuten at folk var mer åpne og villige til å dele følelser med en kunstig intelligent virtuell terapeut enn når de trodde det var en ekte person som styrte den (Lucas et al., 2014). Disse funnene understreker betydningen av menneskelige egenskaper og prosessfaktorer i AI-baserte intervensjoner, og sier noe om at disse menneskelige

egenskapene kan økte individers selvavsløring, også når det sammenlignes med digitale intervensjoner gjennomført med menneskelige moderatører.

Standardiserte mål på terapeutisk allianse. I to av studiene i denne litteraturgjennomgangen ble terapeutisk allianse målt med standardiserte kartleggingsverktøy. Liu et al. (2022) sammenligner sin kontrollgruppe som leste selvhjelpsbok med testgruppen gjennom mål på arbeidsallianse (WAI-SR). He et al. (2022) undersøkte terapeutisk allianse gjennom Working Alliance Questionnaire (WAQ), også med kontrollgruppe som leste selvhjelpsmateriale. Når studiene som målte terapeutisk allianse har benyttet seg av selvhjelpsbøker som kontrollbetingelse, reiser dette spørsmål om hvorvidt det er mulig å sammenligne mål på terapeutisk allianse hos en aktiv og en gruppe uten noe relasjonell kontakt. Siden kontrollgruppene ikke hadde interaksjon, vil det kunne øke risiko for systematisk skjevhet at studien måler arbeidsallianse hos en gruppe som mottok minimal passiv psykoedukasjon, da mål på terapeutisk allianse med stor sannsynlighet ikke vil være like relevante for denne kontrollgruppen. Forskerne i Liu et al. (2022) skriver at terapeutisk allianse, målt ved WAI-SR, i chatbot-testgruppen var betydelig bedre enn i kontrollgruppen med en høy effektstørrelse ($d = 1.85$). Videre tolkes disse resultatene som at bruk av samtalebasert kunstig intelligens kan hjelpe med å bygge en terapeutisk allianse. Dette er resultater som bør leses med forsiktighet, gitt at sammenligningsgrunnlaget er svakt hos den passive kontrollgruppen, som kun leste selvhjelpsmateriale.

Valg av kontrollgrupper

Ingen av studiene som er inkludert i denne litteraturgjennomgangen begrunner sine valg av type kontrollgruppe eksplisitt, og alternative kontrollbetingelser er ikke diskutert. Flere av studiene anmoder imidlertid videre forskning til å inkludere mer aktive kontrolltilnærminger. På motsatt side anmoder Klos et al. (2021) fremtidig forskning til å vurdere bruk av ventelistekontroller dersom det er etisk forsvarlig.

Fire av studiene som er inkludert i denne litteraturgjennomgangen hadde kontrollgrupper som utelukkende skulle lese relevant selvhjelpsmateriale (Fitzpatrick et al., 2017; Fulmer et al., 2018; Klos et al., 2021; Liu et al., 2022), og en studie (He et al., 2022) hadde én kontrollbetingelse som skulle lese en psykoedukativ bok, og én kontrollbetingelse som skulle samhandle med en generell chatbot. Kontrollgruppen i Sadeh-Sharvit et al. (2023) var den eneste som mottok vanlig behandling (TAU) ved en poliklinikk. Terapisesjonene i kontrollgruppen i denne studien ble ikke tatt opp av AI-plattformen, og dette begrenset muligheten for å sammenligne kontrollgruppen og intervensjonsgruppen.

Å inkludere kontrollgrupper som utelukkende mottok selvhjelpsmateriale som intervensjon, kan ha ført til flere begrensninger i de aktuelle studiene. Ifølge Fitzpatrick et al. (2017) begrenser det deres studie ved at objektive kvantitative data om engasjement ikke var tilgjengelige fra kontrollgruppen, og at det derfor ikke var mulig å sammenligne engasjementet mellom de to gruppene på en meningsfull måte. I tillegg, fordi dataene var deidentifiserte, var det ikke mulig å utforske om eventuelle dose-respons-effekter eksisterte. Samtidig bemerkes det i studien at valget av kontrollgruppen kan ha bidratt til den høye frafallsraten. Selvhjelpsmaterialet introduserte heller ikke CBT-spesifikt materiale, så det var ikke mulig å evaluere om den samtalebaserte leveringingen medierte symptomreduksjonen, i stedet for CBT-innholdet som chatboten leverte. Fulmer et al. (2018) bemerker seg i sin studie at deltakerne opplevde en liten økning i symptomer på angst og depresjon, noe som kan skyldes at e-boken ikke var tilstrekkelig som en form for mental helsestøtte. Forskerne understreker at fremtidige studier bør inkludere kontrollbetingelser som tillater en mer direkte sammenligning mellom levering av tjenester som tradisjonell terapi, samt teknologibaserte løsninger.

Som svar på denne oppfordringen introduserte He et al. (2022) i tillegg til en kontrollbetingelse som skulle lese selvhjelpsmateriale, en kontrollgruppe som kun skulle ha

samtaler med en generell chatbot, med bakgrunn i at det ville være enklere å sammenligne resultatene med intervensjonsgruppen. Forskerne i denne studien bemerket også at e-bokintervensjonen manglet omfattende terapiprogrammer, og dermed ble det vanskeligere å sammenligne med de andre aktive gruppene. Forskerne oppfordret til at fremtidige studier bør vurdere alternative aktive kontrolltilnæringer med tydelig dokumentert effekt. Mangelen på objektive atferdsdata i kontrollgruppene begrenset muligheten for direkte sammenligning med behandlingsgruppen, særlig når bevisstryken for intervensjonen i kontrollgruppen var begrenset.

Selv om Fitzpatrick et al. (2017) trekker frem begrensninger ved å benytte seg av selvhjelpsmateriale for kontrollgruppen, trekkes det også frem at styrken til kontrollgruppen kan illustreres ved at individer som avla data i kontrollgruppen opplevde en lignende reduksjon i angstsymptomer som den eksperimentelle gruppen. Dette støtter ifølge forskerne den eksisterende litteraturen, som antyder at minimal passiv psykoedukasjon alene kan redusere symptomer på psykisk belastning (Donker et al., 2009).

Deltakerkarakteristika og symptomnivå hos deltakerne

Flere av studiene som er inkludert i denne gjennomgangen hadde et begrenset antall deltakere, som bidrar til å begrense generaliserbarheten (Fitzpatrick et al., 2017; Fulmer et al., 2018; Sadeh-Sharvit et al., 2023). Majoriteten av studiene undersøker college-studenter, som fører til mangel på deltakervariasjon, og økonomisk status har ikke blitt vurdert i noen av studiene. Den eneste studien som ikke undersøker college-studenter, er Sadeh-Sharvit et al. (2023), som har et utvalg som består av pasienter henvist til en universitetsklinikk, men som til gjengjeld har et svært lavt antall deltakere. De fleste studiene som har rekruttert fra spesifikke studiesteder knyttet til bestemte regioner og land, påpeker at dette er en sentral begrensning for generaliserbarheten av studiene.

Flere av studiene inkluderte en ikke-klinisk populasjon, og hadde ingen krav om grad av symptomer ved baseline. På grunn av rekruttering av deltakere fra et hovedsakelig ikke-klinisk utvalg, var utgangspunktet for symptomtrykk ved depresjon og angst lavt i de fleste studiene. Studiene til Fulmer et al. (2018) og Fitzpatrick et al. (2017) hadde kun krav om at deltakerne selvidentifiserte seg med symptomer på angst og depresjon. Klos et al. (2021) undersøkte også en populasjon hvor det ikke var spesifikke krav om symptomer ved baseline. Liu et al. (2022) hadde et minimumskrav om PHQ-9-score på ni eller høyere, og forskerne i studien til He et al. (2022) krevde gjennomsnittlig poengsum på CSMHSS mellom to og tre. I studien til Sadeh-Sharvit et al. (2023) var deltakerne henvist med en angst eller depresjonslidelse til en poliklinikk, og deltakerne i studien hadde i hovedsak lette til moderate symptomer på depresjon og angst ved starten av studien. Samtlige studier inkludert i denne gjennomgangen oppfordrer fremtidig forskning til å undersøke anvendeligheten og effektiviteten av AI-plattformer for behandling av personer med alvorlige psykiske lidelser og høyere symptomtrykk enn det som er undersøkt til nå.

Av de seks studiene inkludert i denne gjennomgangen, er det altså kun tre studier som undersøker en populasjon som inkluderer personer med moderat grad av angst og depresjon basert på de validerte selvrappoteringskjemaene som er benyttet. At studiene benytter seg av deltakere med varierende symptomnivå ved baseline, og i tillegg benytter seg av forskjellige standardiserte mål på symptomer, gjør det utfordrende å sammenligne resultatene i studiene. Liu et al. (2022) begrenset også terapimålet i sin studie til depresjon, selv om klienter ofte har en kombinasjon av problemer, og påpeker at dette kan begrense generaliserbarheten av funnene.

Systemfeil og innholdsmessige begrensninger

Karakteristika knyttet til design og innhold kan regnes som sentrale begrensninger i flere av studiene. Alle studiene som benyttet seg av fullautomatiserte chatbots trekker frem

systemfeil, begrenset innhold og tekniske problemer som årsaker til redusert generaliserbarhet. For eksempel trekker Fulmer et al. (2018) frem at systemfeil ble uttrykt av flere deltakere under kvalitativ tilbakemelding, og forklares av forskerne som et resultat av begrensede ressurser og forsøk på å beholde all godkjent innhold fra eksperter intakt. I løpet av studien ble endringer i systemet begrenset, og derfor kunne forskningsteamet ikke rapportere feil relatert til naturlig språkbehandling eller misforhold i følelser før studien var avsluttet. På samme måte nevner Liu et al. (2022) at innholdet i deres chatbot var begrenset for å sikre faglig godkjenning og hadde tekniske begrensninger som ikke tok høyde for repetisjonshåndtering og personalisering. Fitzpatrick et al. (2017) mottok også tilbakemeldinger som gikk på systemfeil og innholdsmessige begrensninger, der hovedtemaene var brudd på prosessen, tekniske problemer og problemer med innholdet. Særlig fremtredende var tilbakemeldinger om begrensninger i naturlige samtaler, inkludert manglende forståelse av visse svar, gjentakende samtalesegmenter og forvirring når deltakerne ga uventede svar. Problemer med innholdet involverte hovedsakelig utfordringer knyttet til emoji'er samt interaksjoner og lengden på innholdet. Klos et al. (2021) mottok tilbakemeldinger om at brukerne ikke alltid følte seg forstått, og at brukerne ikke fikk svar på det de spurte om, og He et al. (2022) beskriver at innhold og teknologi var de mest rapporterte negative erfaringene ved intervensjonen. Tilbakemeldinger om systemfeil og innhold kan begrense resultatene, og særlig påvirke resultater knyttet til brukervennlighet, akseptabilitet og frafallsrater.

Kan man egentlig blinde deltakere i slike studier?

Begrepene blindet og åpen referer i forskningssammenheng til måten deltakere og/eller forskere blir informert om behandlingsgruppene i en randomisert kontrollert studie. Deltakerblindet betyr at deltakerne i studien ikke vet hvilken behandlingsgruppe de er i, og minimerer muligheten for at deltakerne påvirkes av kunnskapen om behandlingsgruppen

deres. Forskerblindet betyr at de som gjennomfører studien heller ikke vet hvilken behandlingsgruppe hver deltaker tilhører, og reduserer muligheten for at forskerne ubevisst påvirker studiens resultater eller tolkninger. I en åpen RCT-studie vet både deltakerne og forskerne hvilken behandlingsgruppe hver deltaker tilhører. Bruken av blindede studier bidrar til å redusere systematiske skjevheter i resultatene. Åpne RCT-studier brukes der det er praktisk umulig å blinde deltakerne eller forskerne, for eksempel ved spesielle typer terapier der deltakerne åpenbart vet hvilken behandling de mottar.

Ingen av studiene inkludert i denne gjennomgangen rapporterer at de er deltakerblindet. To av studiene (Fitzpatrick et al., 2017; Liu et al., 2022) oppgir at de er åpne (ikke-blindede) studier, og tre av studiene oppgir ingen informasjon om hvorvidt forskerne eller deltakerne er blindet eller ikke (Fulmer et al., 2018; Klos et al., 2021; Sadeh-Sharvit et al., 2023). Forskerne i studien til He et al. (2022) rapporterer at deres studie er forskerblindet. Behandlingsfordelingen ble i denne studien skjult for deltakere, forskere og de som var involvert i analyse av forsøksdata, da den ble lagret i en kryptert elektronisk filform av flere parter, og avslørt etter fullføring av dataanalysen. Intervensjonen så vel som utfallsmålingen ble gjennomført online, og ingen av forskerne hadde tilgang til deltakernes systemer i løpet av intervensjonsperioden (enkel blind). He et al. (2022) trekker frem at dobbeltblindføring av både forskere og deltakere umulig på grunn av verktøyets egenart og behovet for å opprettholde rekruttering. Dette kan ha ført til subjektiv innflytelse på intervensjonens effektivitet. Det at majoriteten av studiene i denne litteraturgjennomgangen ikke er blindet, gjør at funnene må leses med forsiktighet, fordi deltakere og forskere kan ha vært påvirket av informasjonen de har hatt om de ulike intervensjonene.

Behandlingsmetode og oppfølging

I denne litteraturgjennomgangen peker flere studier i samme retning når det gjelder resultater på symptomreduksjon, noe som indikerer at behandlingsmetoden tilsynelatende har

en effekt allerede i de tidlige stadiene av terapiforløpet. Det reises imidlertid viktige spørsmål om hvorvidt denne effekten stabiliseres over tid, hvilke parametere som blir målt, og om visse karakteristika ved den målte populasjonen påvirker utfallet. For å oppnå en helhetlig forståelse av psykoterapi levert av chatbots, er det nødvendig å analysere både kortsiktige og langsiktige resultater, samtidig som man vurderer variasjoner i effektivitet basert på ulike behandlingsmetoder og pasientpopulasjoner.

En vesentlig faktor som må tas i betraktning i denne litteraturgjennomgangen er som nevnt at de inkluderte studiene ikke er deltakerblindet. Deltakelse i seg selv kan dermed potensielt påvirke symptomer, og det er plausibelt å anta at deltakelse i studien kan føre til at deltakerne bryter med eksisterende atferdsmønstre. Dette kan skyldes flere faktorer, som at deltakerne finner deltakelsen engasjerende eller underholdende. Spesielt kan ung alder og oppvekst i en digital tidsalder spille en rolle i deltakernes tro på intervensjonene, og det kan argumenteres for at selve deltakelsen kan ha en innvirkning på symptomer i seg selv. Deltakelse kan skape endringer i hverdagen og bidra til å bryte mønstre, og gi lignende virkninger som man ser i begynnelsen av tradisjonell terapi der individet blir oppfordret til å utforske nye atferdsmønstre og opplevelser. Det kan tenkes at denne effekten av deltakelse kan ses i sammenheng med virkningene man observerer i startfasen av konvensjonell terapi. For å nærmere undersøke om effekten av behandlingen er høyest i startfasen og deretter stabiliseres, eller om effektene er langvarige, er det essensielt å gjennomføre oppfølgingsstudier innenfor dette forskningsfeltet. Dette vil bidra til å kaste lys over varigheten og dybden av de observerte terapeutiske effektene, samt avklare om deltakelse i seg selv spiller en betydelig rolle i symptomreduksjon.

Begrensninger og styrker ved denne litteraturgjennomgangen

Et strukturert datainnsamlings skjema ble utviklet for å hente relevant informasjon fra de seks inkluderte studiene. Risikoen for systematisk skjevhet ble vurdert med tanke på

studiedesign, utvalgskriterier, utvalgsbias, seleksjonsskjevhet, publiseringskjevhet, og rapporteringsbias. Styrkene ved denne litteraturgjennomgangen inkluderer klare inklusjons- og eksklusjonskriterier, grundig vurdering av risiko for systematisk skjevhet, og bruk av standardiserte utfallsmål. En annen styrke ved denne litteraturgjennomgangen er at alle inkluderte studier var RCT-studier. For å styrke bruk av kunstig intelligens i behandling for psykiske utfordringer, er det viktig at forskningsgrunnlaget er evidensbasert. Et viktig aspekt ved å kalle en behandlingsform for evidensbasert, er at forskningsgrunnlaget består av RCT-studier. Andre styrker ved denne litteraturgjennomgangen er lav grad av variasjon i de standardiserte, validerte utfallsmålene, slik at funnene kan sammenlignes med utgangspunkt innenfor samme forholdsmessige ramme. Det er også en bred geografisk, kulturell og språkmessig variasjon i studiene inkludert i denne gjennomgangen, som bidrar til en bredere ekstern validitet, selv om flere metodologisk styrkede studier må gjennomføres.

Det er betydelige begrensninger ved denne litteraturgjennomgangen, deriblant metodologiske og innholdsmessige begrensninger, og risiko for systematisk skjevhet. Direkte sammenligning av resultater er vanskelig grunnet variasjonen i studiene, som begrenser muligheten til å fastslå virkningen og de underliggende mekanismene bak virkningen av kunstig intelligens for reduksjon av angst- og depressive symptomer. Variasjonen i blant annet intervensjonsmetoder, kontrollbetingelser, frafall og varighet av intervensjonene, i tillegg til variasjon i rapportering av resultater, er noe som gjør det utfordrende å trekke generelle konklusjoner.

Få studier er inkludert i denne litteraturgjennomgangen, og funnene er derfor for små til at det kan trekkes slutninger basert på funnene. I tillegg må det tas forbehold om at kunstig intelligens innenfor mental helsebehandling er et relativt nytt felt, og det er ikke etablert et fullstendig og anerkjent begrepsapparat eller teoretisk rammeverk for å undersøke og forstå disse problemstillingene. Det må derfor tas forbehold om at litteratursøket som er

gjennomført kan ha ekskludert studier som var relevante for problemstillingen, på tross av at søkeord var nøye vurdert opp imot tidligere litteratur, problemstilling og kontekst. Inklusjons- og eksklusjonskriterier er nøye spesifisert for populasjon, studiedesign, intervensjoner, utfallsmål, språk, og publikasjonsår, men det kan likevel være en risiko for at relevante studier ikke er oppdaget. Det er også en risiko for publiseringskjevhet og usikkerhet knyttet til risikovurdering av systematisk skjevhet som må tas høyde for i vurderingen av resultatene av denne litteraturgjennomgangen.

Implikasjoner og forslag til videre forskning

Resultatene av denne systematiske gjennomgangen gir innsikt i effektiviteten til automatiserte psykologiske chatbots og AI-støtteverktøy som behandlingsalternativer for depresjon og angst. Funnenes relevans ses også i sammenheng med og understrekes gjennom psykologisk teori knyttet til antropomorfisme og psykologisk avstand, som er foreslåtte teoretiske rammeverk for bruk av kunstig intelligens innen psykologi.

Litteraturgjennomgangens resultater vil kunne ha implikasjoner for videre forskning og for klinisk praksis. Sammenligninger mellom studiene kan avdekke mulige forskjeller i effektstørrelser og akseptabilitet blant ulike AI-intervensjoner, og gjennomgangen kan informere fremtidig forskning og praksis når det gjelder bruk av teknologi i psykisk helsebehandling.

Det fremgår et behov for ytterligere studier med større utvalgsstørrelse og standardiserte metoder for å validere funnene. Særlig mangler det studier som fokuserer på en ren klinisk populasjon, med spesifikke krav til symptomnivåer ved baseline. Studier som undersøker rene kliniske populasjoner, vil kunne bidra til en utvikling av behandlingstilbud, som kan bidra til å sikre flere mennesker tilgang til behandling som er basert på et robust evidensgrunnlag. Det er reist flere bekymringer knyttet til manglende etiske vurderinger av chatbots og AI-plattformer, og fremtidig søkelys på spesifikke grupper vil kunne bidra til å

sikre en mer etisk praksis. Fremtidige studier også bør presisere hvilke spesifikke angst- og depressive symptomer som undersøkes, og vurdere eventuelle variasjoner i effekten på disse symptomene. Dette vil ikke bare gjøre det enklere å sammenligne rapporterte resultater mellom studier, men det vil også kunne sikre skreddersydd behandling knyttet til spesifikke symptomer. Videre forskning bør også utforske hvordan chatbots med menneskelige egenskaper sammenlignes med faktiske mennesker, for å sikre nøyaktige konklusjoner og objektiv rapportering av resultater, spesielt med tanke på å forstå betydningen av menneskelige egenskaper i behandling med kunstig intelligens.

Metodologisk kvalitet kan forbedres på flere områder. Langsiktige studier med oppfølgingsdata er nødvendige for å undersøke varige effekter av intervensjoner og styrke funnene angående virkning og årsakssammenhenger. Implementering av blinding ved utfallsvurderinger bør vurderes for å redusere subjektivitet og skjevhet, og forskningsprotokoller bør tydelig inkorporere retningslinjer for blinding, spesielt ved evaluering av terapeutiske effekter. Forskerne bør også vurdere å begrense selvrapporing for å minimere risikoen for skjevhet i effektstørrelser, og ytterligere standardisering av kliniske vurderinger knyttet til utfallsmål vil bidra til mer konsistente og pålitelige resultater. For å styrke statistisk kraft og generaliserbarhet, bør fremtidige studier vurdere økning i utvalgsstørrelse og intervensjonsperiode. Inkludering av aktive kontrollgrupper eller ventelistekontroller bør også vurderes, da dette kan gi verdifull innsikt. Samlet sett kan disse tiltakene bidra til å forbedre evidensgrunnet og øke forståelsen av bruken av kunstig intelligens i behandling av symptomer på angst og depresjon. Dette vil samtidig kunne ha viktige implikasjoner for implementering av effektive behandlingstilbud.

Konklusjon

Denne systematiske litteraturgjennomgangen indikerer at psykologisk terapi basert på kunstig intelligens kan være effektivt for å redusere symptomer på angst og depresjon. Chatbots og AI-plattformer viser seg særlig å være nyttige i behandlingen av depresjon, selv om ytterligere forskning er nødvendig for å forstå og forbedre deres effektivitet, mens effektene på angst varierer. Studiene viser variasjon i rapportering av resultater, frafallsrater og metoder, og det er viktig å utøve forsiktighet ved tolkning av funnene. Én studie fant ingen signifikant effekt på depresjon, mens andre observerer betydelige reduksjoner. Forskjeller i tolkning av selvrapporteringsskjemaer, manglende deltakerblindhet og variasjon i intervensjonsmetoder er potensielle kilder til usikkerhet og begrensninger i resultatene.

Gjennomgangen utforsker også hvordan terapi basert på kunstig intelligens virker, med psykologisk distanse og antropomorfisme som foreslåtte underliggende mekanismer. Resultatene tyder på at chatbots kan skape opplevelser som ligner menneskelig interaksjon, spesielt når det gjelder å håndtere depresjon. Antropomorfisme, psykologisk avstand og terapeutisk allianse blir identifisert som mulige mekanismer som påvirker intervensjonenes effekt. Frafallsrater varierer betydelig, men terapichatbots viser høyere overholdelsesrate sammenlignet med tidligere internettbaserte psykologiske intervensjoner, og sammenlignet med kontrollgruppene i studiene.

Begrensninger i denne gjennomgangen inkluderer metodologiske utfordringer, heterogenitet i studiedesign og populasjoner, samt risiko for systematisk skjevhet. Det understrekes også at kunstig intelligens i mental helsebehandling er et relativt nytt felt, som fører til et begrenset antall studier, og videre forskning er nødvendig for å etablere robuste funn og forståelse av virkningen. Fremtidig forskning oppfordres til å styrke metodologisk kvalitet, utforske varigheten av effektene ved å gjennomføre studier med lengre

intervensjonsperioder og samle inn oppfølgingsdata. En utforskning og sammenligning av ulike former for psykologiske intervensjoner basert på kunstig intelligens, samt sammenligning med intervensjoner ledet av menneskelige moderatorer, vil ytterligere berike feltet.

Denne systematiske gjennomgangen vil kunne gi viktig innsikt i potensialet for automatiserte chatbots og AI-plattformer som effektive verktøy for å redusere depressive og angstsymptomer. Det vil også bidra til å forstå akseptabilitet og metodologiske forbedringspotensial knyttet til slike intervensjoner. Bruk av Chatbots og AI-plattformer kan være en kostnadseffektiv måte å redusere symptomer på angst og depresjon, samtidig som uveiledede intervensjoner basert på teknologi kan bidra til å redusere behandlingsutilgjengelighet, og støtte de som lider av milde og moderate symptomer på depresjon og angst.

Referanser

- Abd-alrazaq, A. A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P., & Househ, M. (2019). An overview of the features of chatbots in mental health: A scoping review. *International Journal of Medical Informatics*, *132*, 103978. <https://doi.org/https://doi.org/10.1016/j.ijmedinf.2019.103978>
- Abrams, E. M., Akombi, B., Alam, S., Alcalde-Rabanal, J. E., Allebeck, P., Amini-Rarani, M., Atnafu, D. D., Ausloos, M., Ayanore, M. A., Azari, S., Azene, Z. N., Babae, E., Bante, A. B., Bisanzio, D., Bohluli, M., Breitborde, N. J. K., Briant, P. S., Castelpietra, G., Cunningham, M., . . . Yip, P. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet (British Edition)*, *396*(10258), 1204-1222. [https://doi.org/10.1016/S0140-6736\(20\)30925-9](https://doi.org/10.1016/S0140-6736(20)30925-9)
- Ahmed, A., Ali, N., Aziz, S., Abd-Alrazaq, A. A., Hassan, A., Khalifa, M., Elhusein, B., Ahmed, M., Ahmed, M. A. S., & Househ, M. (2021). A review of mobile chatbot apps for anxiety and depression and their self-care features. *Computer Methods and Programs in Biomedicine Update*, *1*, 100012. <https://doi.org/10.1016/j.cmpbup.2021.100012>
- Ahn, J., Kim, J., & Sung, Y. (2021). AI-powered recommendations: the roles of perceived similarity and psychological distance on persuasion. *International Journal of Advertising*, *40*(8), 1366-1384. <https://doi.org/10.1080/02650487.2021.1982529>
- Ali, M. R., Razavi, Z., Abdullah Al Mamun, Langevin, R., Kane, B., Rawassizadeh, R., Schubert, L., & M Ehsan Hoque. (2020). A Virtual Conversational Agent for Teens with Autism: Experimental Results and Design Lessons. *arXiv.org*. <https://doi.org/10.1145/3383652.3423900>

American Psychiatric Association, D., & American Psychiatric Association. (2013).

Diagnostic and statistical manual of mental disorders: DSM-IV 5(5). Washington DC: American Psychiatric Association.

Andersson, G. (2016). Internet-Delivered Psychological Treatments. *Annual Review of Clinical Psychology*, *12*(1), 157-179. <https://doi.org/10.1146/annurev-clinpsy-021815-093006>

Antoniadi, A. M., Du, Y., Guendouz, Y., Wei, L., Mazo, C., Becker, B. A., & Mooney, C. (2021). Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review. *Applied Sciences*, *11*(11), 5088. <https://doi.org/10.3390/app11115088>

Balcetis, E., & Dunning, D. (2010). Wishful Seeing: More Desired Objects Are Seen as Closer. *Psychological Science*, *21*(1), 147-152. <https://doi.org/10.1177/0956797609356283>

Bickmore, T., Gruber, A., & Picard, R. (2005). Establishing the computer–patient working alliance in automated health behavior change interventions. *Patient Education and Counseling*, *59*(1), 21-30. <https://doi.org/https://doi.org/10.1016/j.pec.2004.09.008>

Borghouts, J., Eikey, E., Mark, G., De Leon, C., Schueller, S. M., Schneider, M., Stadnick, N., Zheng, K., Mukamel, D., & Sorkin, D. H. (2021). Barriers to and Facilitators of User Engagement With Digital Mental Health Interventions: Systematic Review [Review]. *Journal of Medical Internet Research*, *23*(3), e24387. <https://doi.org/10.2196/24387>

Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., Parks, A. C., & Zilca, R. (2021). Artificially intelligent chatbots in digital mental health interventions: a review. *Expert Review of Medical Devices*, *18*, 37-49. <https://doi.org/10.1080/17434440.2021.2013200>

- Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: A review of meta-analyses. *Clinical Psychology Review, 26*(1), 17-31. <https://doi.org/https://doi.org/10.1016/j.cpr.2005.07.003>
- Christensen, H. (2007). Computerised therapy for psychiatric disorders. *The Lancet, 370*(9582), 112-113. [https://doi.org/https://doi.org/10.1016/S0140-6736\(07\)61065-4](https://doi.org/https://doi.org/10.1016/S0140-6736(07)61065-4)
- Chung, S. I., & Han, K.-H. (2022). Consumer perception of Chatbots and Purchase Intentions: Anthropomorphism and Conversational Relevance. *International Journal of Advanced Culture Technology, 10*(1), 211-229.
- Clement, S., Schauman, O., Graham, T., Maggioni, F., Evans-Lacko, S., Bezborodovs, N., Morgan, C., Rüsch, N., Brown, J. S. L., & Thornicroft, G. (2015). What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. *Psychological Medicine, 45*(1), 11-27. <https://doi.org/10.1017/S0033291714000129>
- Cuijpers, P., Reijnders, M., & Huibers, M. J. (2019). The Role of Common Factors in Psychotherapy Outcomes. *Annual Review of Clinical Psychology, 15*(1), 207–231. <https://doi.org/10.1146/annurev-clinpsy-050718-095424>
- D'Alfonso, S., Santesteban-Echarri, O., Rice, S., Wadley, G., Lederman, R., Miles, C., Gleeson, J., & Alvarez-Jimenez, M. (2017). Artificial intelligence-assisted online social therapy for youth mental health. *Frontiers in Psychology, 8*, 796. <https://doi.org/10.3389/fpsyg.2017.00796>
- Danieli, M., Ciulli, T., Mousavi, S. M., Silvestri, G., Barbato, S., Di Natale, L., & Riccardi, G. (2022). Assessing the Impact of Conversational Artificial Intelligence in the Treatment of Stress and Anxiety in Aging Adults: Randomized Controlled Trial. *Jmir Mental Health, 9*(9), e38067. <https://doi.org/10.2196/38067>

- Darcy, A., Daniels, J., Salinger, D., Wicks, P., & Robinson, A. (2021). Evidence of Human-Level Bonds Established With a Digital Conversational Agent: Cross-sectional, Retrospective Observational Study. *JMIR Formative Research*, 5(5), e27868. <https://doi.org/10.2196/27868>
- DeAngelis, T. (2021). Depression and anxiety escalate during COVID. *Monitor on Psychology*, 52(8). <https://www.apa.org/monitor/2021/11/numbers-depression-anxiety>
- Donker, T., Griffiths, K. M., Cuijpers, P., & Christensen, H. (2009). Psychoeducation for depression, anxiety and psychological distress: a meta-analysis. *BMC Medicine*, 7(1), 79. <https://doi.org/10.1186/1741-7015-7-79>
- Donkin, L., Christensen, H., Naismith, S. L., Neal, B., Hickie, I. B., & Glozier, N. (2011). A Systematic Review of the Impact of Adherence on the Effectiveness of e-Therapies [Original Paper]. *Journal of Medical Internet Research*, 13(3), e52. <https://doi.org/10.2196/jmir.1772>
- Dosovitsky, G., & Bunge, E. L. (2021). Bonding With Bot: User Feedback on a Chatbot for Social Isolation [Original Research]. *Frontiers in Digital Health*, 3. <https://doi.org/10.3389/fdgth.2021.735053>
- Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and autonomous systems*, 42(3-4), 177-190. [https://doi.org/10.1016/S0921-8890\(02\)00374-3](https://doi.org/10.1016/S0921-8890(02)00374-3)
- Ebert, D. D., Harrer, M., Apolinario-Hagen, J., & Baumeister, H. (2019). Digital Interventions for Mental Disorders: Key Features, Efficacy, and Potential for Artificial Intelligence Applications. In Y. K. Kim (Ed.), *Frontiers in Psychiatry: Artificial Intelligence, Precision Medicine, and Other Paradigm Shifts*, vol 1192. 583-627. https://doi.org/10.1007/978-981-32-9721-0_29

- Eells, T. D., Barrett, M. S., Wright, J. H., & Thase, M. (2014). Computer-assisted cognitive-behavior therapy for depression. *Psychotherapy (Chic)*, *51*(2), 191-197.
<https://doi.org/10.1037/a0032406>
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychological review*, *114*(4), 864-886.
<https://doi.org/10.1037/0033-295X.114.4.864>
- Fairburn, C. G., & Patel, V. (2017). The impact of digital technology on psychological treatments and their dissemination. *Behaviour Research and Therapy*, *88*, 19-25.
<https://doi.org/https://doi.org/10.1016/j.brat.2016.08.012>
- Fiske, A., Henningsen, P., & Buyx, A. (2019). Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy. *Journal of Medical Internet Research*, *21*(5), e13216.
<https://doi.org/10.2196/13216>
- Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *Jmir Mental Health*, *4*(2), e19. <https://doi.org/10.2196/mental.7785>
- Fortune Business Insights. (2023). Artificial intelligence (AI) market size, share and industry analysis by component (hardware, software, services), by technology (computer vision, machine learning, natural language processing, others), by industry vertical (BFSI, healthcare, manufacturing, retail, IT & telecom, government, others) and regional forecast. <https://www.fortunebusinessinsights.com/industry-reports/artificial-intelligence-market-100114#companiesstudies>
- Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., & Rauws, M. (2018). Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety:

- Randomized Controlled Trial. *Jmir Mental Health*, 5(4), e64.
<https://doi.org/10.2196/mental.9782>
- Gaffney, H., Mansell, W., & Tai, S. (2019). Conversational Agents in the Treatment of Mental Health Problems: Mixed-Method Systematic Review. *Jmir Mental Health*, 6(10), e14166. <https://doi.org/10.2196/14166>
- Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, 97, 304-316. <https://doi.org/10.1016/j.chb.2019.01.020>
- Goethals, G. R., & Nelson, R. E. (1973). Similarity in the influence process: The belief-value distinction. *Journal of Personality and Social Psychology*, 25(1), 117-122.
<https://doi.org/10.1037/h0034266>
- Greer, S., Ramo, D., Chang, Y.-J., Fu, M., Moskowitz, J., & Haritatos, J. (2019). Use of the Chatbot "Vivibot" to Deliver Positive Psychology Skills and Promote Well-Being Among Young People After Cancer Treatment: Randomized Controlled Feasibility Trial. *JMIR mHealth and uHealth*, 7(10), e15018–e15018.
<https://doi.org/10.2196/15018>
- Han, J. J., Gershoff, A. D., Kirmani, A., & Dalton, A. (2018). When Good Things Feel Closer and Bad Things Feel Farther: The Role of Perceived Control on Psychological Distance Perception. *Journal of consumer psychology*, 28(4), 629-643.
<https://doi.org/10.1002/jcpy.1034>
- Hauser-Ulrich, S., Künzli, H., Meier-Peterhans, D., & Kowatsch, T. (2020). A Smartphone-Based Health Care Chatbot to Promote Self-Management of Chronic Pain (SELMA): Pilot Randomized Controlled Trial. *Jmir Mhealth and Uhealth*, 8(4), e15806.
<https://doi.org/10.2196/15806>

- He, Y., Yang, L., Zhu, X., Wu, B., Zhang, S., Qian, C., & Tian, T. (2022). Mental Health Chatbot for Young Adults With Depressive Symptoms During the COVID-19 Pandemic: Single-Blind, Three-Arm Randomized Controlled Trial. *Journal of Medical Internet Research, 24(11)*, e40719. <https://doi.org/10.2196/40719>
- Ienca, M., Wangmo, T., Jotterand, F., Kressig, R. W., & Elger, B. (2018). Ethical Design of Intelligent Assistive Technologies for Dementia: A Descriptive Review. *Science and Engineering Ethics, 24(4)*, 1035-1055. <https://doi.org/10.1007/s11948-017-9976-1>
- Inkster, B., Sarda, S., & Subramanian, V. (2018). An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *Jmir Mhealth and Uhealth, 6(11)*, e12106. <https://doi.org/10.2196/12106>
- Kaiser, J., Hanschmidt, F., & Kersting, A. (2021). The association between therapeutic alliance and outcome in internet-based psychological interventions: A meta-analysis. *Computers in Human Behavior, 114*, 106512. <https://doi.org/https://doi.org/10.1016/j.chb.2020.106512>
- Kazdin, A. E., & Blase, S. L. (2011). Rebooting Psychotherapy Research and Practice to Reduce the Burden of Mental Illness. *Perspectives on Psychological Science, 6(1)*, 21-37. <https://doi.org/10.1177/1745691610393527>
- Kellogg, K. C., & Sadeh-Sharvit, S. (2022). Pragmatic AI-augmentation in mental healthcare: Key technologies, potential benefits, and real-world challenges and solutions for frontline clinicians. *Frontiers in Psychiatry, 13*. <https://doi.org/10.3389/fpsyt.2022.990370>
- Klos, M. C., Escoredo, M., Joerin, A., Lemos, V. N., Rauws, M., & Bunge, E. L. (2021). Artificial Intelligence-Based Chatbot for Anxiety and Depression in University

- Students: Pilot Randomized Controlled Trial. *Jmir Formative Research*, 5(8), e20678.
<https://doi.org/10.2196/20678>
- Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y. S., & Coiera, E. (2018). Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, 25(9), 1248-1258. <https://doi.org/10.1093/jamia/ocy072>
- Lattie, E. G., Adkins, E. C., Winkvist, N., Stiles-Shields, C., Wafford, Q. E., & Graham, A. K. (2019). Digital Mental Health Interventions for Depression, Anxiety, and Enhancement of Psychological Well-Being Among College Students: Systematic Review [Original Paper]. *Journal of Medical Internet Research*, 21(7), <https://doi.org/10.2196/12869>
- Li, X. G., & Sung, Y. J. (2021). Anthropomorphism brings us closer: The mediating role of psychological distance in User-AI assistant interactions. *Computers in Human Behavior*, 118, 106680. <https://doi.org/10.1016/j.chb.2021.106680>
- Liberman, N., & Trope, Y. (1998). The Role of Feasibility and Desirability Considerations in Near and Distant Future Decisions. *Journal of Personality and Social Psychology*, 75(1), 5–18. <https://doi.org/10.1037/0022-3514.75.1.5>
- Liu, H., Peng, H. M., Song, X. Y., Xu, C. Z., & Zhang, M. (2022). Using AI chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness. *Internet Interventions-the Application of Information Technology in Mental and Behavioural Health*, 27, 100495.
<https://doi.org/10.1016/j.invent.2022.100495>
- Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37, 94-100.
<https://doi.org/https://doi.org/10.1016/j.chb.2014.04.043>

- Ma, T., Sharifi, H., & Chattopadhyay, D. (2019). Virtual Humans in Health-Related Interventions: A Meta-Analysis. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow, Scotland Uk.
<https://doi.org/10.1145/3290607.3312853>
- McTear, M., Callejas, Z., & Griol, D. (2016). The Conversational Interface: Talking to smart devices. *Springer International Publishing AG*.
<https://doi.org/10.1007/978-3-319-32967-3>
- Merriam-Webster Dictionary. (u.å). Artificial intelligence. *Merriam-Webster.com dictionary*.
<https://www.merriam-webster.com/dictionary/artificial%20intelligence>
- Miner, A. S., Milstein, A., & Hancock, J. T. (2017). Talking to Machines About Personal Mental Health Problems. *JAMA : the Journal of the American Medical Association*, 318(13), 1217–1218. <https://doi.org/10.1001/jama.2017.14151>
- Mithen, S., & Boyer, P. (1996). Anthropomorphism and the Evolution of Cognition. *The Journal of the Royal Anthropological Institute*, 2(4), 717–721.
- Mujeeb, S., Hafeez, M., & Arshad, T. (2017). Aquabot: A Diagnostic Chatbot for Achluophobia and Autism. *International Journal of Advanced Computer Science & Applications*, 8(9). <https://doi.org/10.14569/IJACSA.2017.080930>
- Nowak, K. L., & Rauh, C. (2005). The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *Journal of Computer-Mediated Communication*, 11(1), 153-178.
<https://doi.org/10.1111/j.1083-6101.2006.tb00308.x>
- Oladeji, B. D., & Gureje, O. (2016). Brain drain: a challenge to global mental health. *BJPsych international*, 13(3), 61-63. <https://doi.org/10.1192/S2056474000001240>
- Park, G., Chung, J. Y., & Lee, S. Y. (2023). Human vs. machine-like representation in chatbot mental health counseling: the serial mediation of psychological distance and

- trust on compliance intention. *Current Psychology (New Brunswick, N.J.)*, 1–12.
<https://doi.org/10.1007/s12144-023-04653-7>
- Prochaska, J. J., Vogel, E. A., Chieng, A., Kendra, M., Baiocchi, M., Pajarito, S., & Robinson, A. (2021). A therapeutic relational agent for reducing problematic substance use (Woebot): development and usability study. *Journal of Medical Internet Research*, 23(3), e24850. <https://doi.org/10.2196/24850>
- Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people. *Cambridge, UK*, 10(10).
- Richards, D., & Richardson, T. (2012). Computer-based psychological treatments for depression: A systematic review and meta-analysis. *Clinical Psychology Review*, 32(4), 329-342. <https://doi.org/https://doi.org/10.1016/j.cpr.2012.02.004>
- Rost, T., Stein, J., Löbner, M., Kersting, A., Luck-Sikorski, C., & Riedel-Heller, S. G. (2017). User Acceptance of Computerized Cognitive Behavioral Therapy for Depression: Systematic Review. *Journal of Medical Internet Research*, 19(9), e309.
<https://doi.org/10.2196/jmir.7662>
- Sadeh-Sharvit, S., Del Camp, T., Horton, S. E., Hefner, J. D., Berry, J. M., Grossman, E., & Hollon, S. D. (2023). Effects of an Artificial Intelligence Platform for Behavioral Interventions on Depression and Anxiety Symptoms: Randomized Clinical Trial. *Journal of Medical Internet Research*, 25, e46781. <https://doi.org/10.2196/46781>
- Santomauro, D. F., Mantilla Herrera, A. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., Abbafati, C., Adolph, C., Amlag, J. O., Aravkin, A. Y., Bang-Jensen, B. L., Bertolacci, G. J., Bloom, S. S., Castellano, R., Castro, E., Chakrabarti, S., Chattopadhyay, J., Cogen, R. M., Collins, J. K., . . . Ferrari, A. J. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and

- territories in 2020 due to the COVID-19 pandemic. *The Lancet*, 398(10312), 1700-1712. [https://doi.org/10.1016/S0140-6736\(21\)02143-7](https://doi.org/10.1016/S0140-6736(21)02143-7)
- Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J. W., Patel, V., & Silove, D. (2014). The global prevalence of common mental disorders: a systematic review and meta-analysis 1980–2013. *International Journal of Epidemiology*, 43(2), 476-493. <https://doi.org/10.1093/ije/dyu038>
- Tanaka, K., Hayakawa, M., Noda, C., Nakamura, A., & Akiyama, C. (2022). Effects of artificial intelligence aibo intervention on alleviating distress and fear in children. *Child and Adolescent Psychiatry and Mental Health*, 16(1), 87. <https://doi.org/10.1186/s13034-022-00519-1>
- Torous, J., Chan, S. R., Yee-Marie Tan, S., Behrens, J., Mathew, I., Conrad, E. J., Hinton, L., Yellowlees, P., & Keshavan, M. (2014). Patient Smartphone Ownership and Interest in Mobile Apps to Monitor Symptoms of Mental Health Conditions: A Survey in Four Geographically Distinct Psychiatric Clinics. *Jmir Mental Health*, 1(1), e5. <https://doi.org/10.2196/mental.4004>
- Torous, J., Staples, P., Shanahan, M., Lin, C., Peck, P., Keshavan, M., & Onnela, J. P. (2015). Utilizing a Personal Smartphone Custom App to Assess the Patient Health Questionnaire-9 (PHQ-9) Depressive Symptoms in Patients With Major Depressive Disorder. *Jmir Mental Health*, 2(1), e8. <https://doi.org/10.2196/mental.3889>
- Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological review*, 117(2), 440-463. <https://doi.org/10.1037/a0018963>
- Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *The Canadian Journal of Psychiatry*, 64(7), 456-464. <https://doi.org/10.1177/0706743719828977>

- Van Boven, L., Kane, J., McGraw, A. P., & Dale, J. (2010). Feeling close: Emotional intensity reduces perceived psychological distance. *Journal of Personality and Social Psychology, 98*(6), 872-885. <https://doi.org/10.1037/a0019262>
- Vermani, M., Marcus, M., & Katzman, M. A. (2011). Rates of detection of mood and anxiety disorders in primary care: a descriptive, cross-sectional study. *Prim Care Companion CNS Disord, 13*(2). <https://doi.org/10.4088/PCC.10m01013>
- Wu, T., Jia, X., Shi, H., Niu, J., Yin, X., Xie, J., & Wang, X. (2021). Prevalence of mental health problems during the COVID-19 pandemic: A systematic review and meta-analysis. *Journal of Affective Disorders, 281*, 91-98. <https://doi.org/https://doi.org/10.1016/j.jad.2020.11.117>
- Xiang, Y.T., Yang, Y., Li, W., Zhang, L., Zhang, Q., Cheung, T., & Ng, C. H. (2020). Timely mental health care for the 2019 novel coronavirus outbreak is urgently needed. *The Lancet Psychiatry, 7*(3), 228-229. [https://doi.org/https://doi.org/10.1016/S2215-0366\(20\)30046-8](https://doi.org/https://doi.org/10.1016/S2215-0366(20)30046-8)
- Xu, B., & Zhuang, Z. (2022). Survey on psychotherapy chatbots. *Concurrency and Computation: Practice and Experience, 34*(7), e6170. <https://doi.org/https://doi.org/10.1002/cpe.6170>
- Xu, C., Xu, Y., Xu, S., Zhang, Q., Liu, X., Shao, Y., Xu, X., Peng, L., & Li, M. (2020). Cognitive Reappraisal and the Association Between Perceived Stress and Anxiety Symptoms in COVID-19 Isolated People [Original Research]. *Frontiers in Psychiatry, 11*. <https://doi.org/10.3389/fpsy.2020.00858>
- Yan, W. J., Ruan, Q. N., & Jiang, K. (2023). Challenges for Artificial Intelligence in Recognizing Mental Disorders. *Diagnostics, 13*(1), 2. <https://doi.org/10.3390/diagnostics13010002>

Yasavur, U., Lisetti, C., & Rishe, N. (2014). Let's talk! speaking virtual counselor offers you a brief intervention. *Journal on Multimodal User Interfaces*, 8(4), 381-398.

<https://doi.org/10.1007/s12193-014-0169-9>