

UNIVERZITA KARLOVA
KATOLICKÁ TEOLOGICKÁ FAKULTA

Katedra systematické teologie a filosofie

Václav Grepl

**Etické problémy při užití vybraných
prostředků umělé inteligence pro
zvládnání úzkosti a deprese**

Diplomová práce

Vedoucí práce: MUDr. Jan Fošum, Dr. theol.

Praha 2023

Prohlášení

1. Prohlašuji, že jsem předkládanou práci zpracoval samostatně a použil jen uvedené prameny a literaturu.
2. Prohlašuji, že práce nebyla využita k získání jiného titulu.
3. Souhlasím s tím, aby práce byla zpřístupněna pro studijní a výzkumné účely.

V Praze 31.5. 2022

Václav Grepl

Bibliografická citace

Etické problémy při užití vybraných prostředků umělé inteligence pro zvládání úzkosti [rukopis]: diplomová práce / Václav Grepl ; vedoucí práce: MUDr. Jan Fošum, Dr. theol. Praha, 2023. s. 80.

Anotace

Diplomová práce se zabývá možnými etickými problémy při užití umělé inteligence jako podpůrného terapeutického nástroje pro zvládání úzkostných stavů. Rozvoj v oblasti umělé inteligence má vzrůstající klinický význam pro její využití v oblasti péče o duševní zdraví. Systémy založené na umělé inteligenci často přebírají některé podpůrné úkoly i v oblasti terapeutických intervencí, které dříve poskytovali kliničtí psychologové. Jedním z hlavních nástrojů psychoterapie je psychoterapeutický rozhovor vedený mezi pacientem a terapeutem. Rozhovor má nezastupitelný význam pro vytvoření terapeutického vztahu, stanovení diagnózy a vlastní terapeutický proces. Tradiční model psychoterapie se rozšiřuje, protože část péče o duševní zdraví začínají v některých technologicky vyspělých zemích nyní poskytovat z kapacitních důvodů i konverzační systémy, vedoucí prostřednictvím umělé inteligence terapeutický rozhovor s příjemcem terapie. Pro eticky odpovědné klinické využívání umělé inteligence je třeba identifikovat možné etické důsledky jejího narůstajícího využívání v oblasti péče o duševní zdraví a vytvořit podklady pro rozhodování o dynamice a mezích jejího využívání, stejně jako dodržovat zásady vývoje umělé inteligence konformní s etickými principy. Cílem diplomové práce je identifikovat a popsat hlavní etické problémy a konflikty při využívání vybraných systémů konverzační umělé inteligence Woebot, Youper a Replika, určených pro podpůrnou terapeutickou pomoc při zvládání úzkostných stavů, a popsat etická rizika předpokládaného budoucího rozšíření používání prostředků umělé inteligence při péči o duševní zdraví.

Klíčová slova

Psychoterapie, úzkostné stavy, psychoterapeutický rozhovor, etické riziko, autonomie, beneficence, maleficence, spravedlnost, umělá inteligence, konverzační robot.

Abstract

The diploma thesis deals with possible ethical problems in the use of artificial intelligence as a supportive therapeutic tool for managing anxiety states. Developments in the field of artificial intelligence are of increasing clinical importance for its use in the field of mental health care. Systems based on artificial intelligence often take over some support tasks even in the field of therapeutic interventions that were previously provided by clinical psychologists. One of the main tools of psychotherapy is the psychotherapeutic interview conducted between the patient and the therapist. The interview is important for creating a therapeutic relationship, determining a diagnosis and for the therapeutic process itself. The traditional model of psychotherapy is expanding, as part of mental health care is now starting to be provided, partly due to capacity reasons, by artificial intelligence conversational systems leading a therapeutic conversation with the recipient of the therapy. For the ethically responsible clinical use of artificial intelligence, it is necessary to identify the possible ethical consequences of the growing use of artificial intelligence in the field of mental health care and to create the basis for both decision-making about the dynamics and limits of its use, as well as the principles of its development ensuring the creation of such an artificial intelligence tool that will be conforming to ethical principles. The aim of the diploma thesis is to identify and describe the main ethical problems and conflicts in the use of selected conversational artificial intelligence systems of Woebot, Youper and Replika, which are intended for supportive therapeutic help in managing anxiety states, and to describe the ethical risks of the expected future expansion of the use of artificial intelligence tools in the care of mental health.

Keywords

Psychotherapy, anxiety, psychotherapeutic interview, ethical risk, autonomy, beneficence, maleficence, equity, artificial intelligence, chatbot.

Počet znaků (včetně mezer): <151.302>

Poděkování

Tímto chci poděkovat vedoucímu mé diplomové práce MUDr. Janu Fošumovi, Dr.theol. za konzultace při výběru tématu diplomové práce, za jeho cenné připomínky k textu a odborné vedení. Rovněž bych chtěl poděkovat mé ženě Ireně za chápající porozumění a bratrovi Vítězslavovi za účinnou podporu.

Obsah

Obsah.....	6
Úvod.....	8
Obecné uvedení tématu – předmět práce	8
Cíl práce a výzkumná otázka.....	8
Aktuálnost tématu.....	9
Použitá metoda	10
Struktura práce	10
1. Základní pojmy a koncepty	11
1.1. Umělá inteligence	11
1.2. Psychoterapie úzkostných poruch a úzkostných stavů	14
1.3. Etika umělé inteligence v psychoterapii	16
2. Popis užitých metod	18
3. Etika psychoterapie při využití umělé inteligence.....	19
3.1. Etika psychoterapie	19
3.2. Etika umělé inteligence v perspektivě bioetiky	23
3.3. Etika umělé inteligence v psychoterapii	25
4. Koncept terapie úzkostných poruch a stavů	32
4.1. Klinický obraz úzkosti	32
4.2. Prevalence úzkostné poruchy.....	33
4.3. Etiologie úzkostných stavů a úzkostných poruch.....	33
4.4. Léčba úzkostných stavů a úzkostných poruch	34
4.5. Metody léčby úzkostných stavů a úzkostných poruch.....	35
5. Etická reflexe specifických konverzačních nástrojů	37
5.1. Nástroje pro podporu péče o duševní zdraví	37

5.2.	Popis vybraných chatbotů.....	41
5.3.	Etické problémy a výzvy při vývoji a užití chatbotů.....	45
5.4.	Limity prostředků péče o duševní zdraví mající etický přesah ..	49
5.5.	Limity chatbotů z perspektivy etiky dialogického personalismu	54
6.	Prevence etických problémů umělé inteligence	56
6.1.	Syntéza etických principů.....	56
6.2.	Etické výzvy pro budoucí vývoj a užití umělé inteligence.....	57
	Závěr.....	70
	Seznam použitých zkratk.....	73
	Seznam použité literatury a pramenů	74

Úvod

Obecné uvedení tématu – předmět práce

Umělá inteligence je často popisována jako čtvrtá průmyslová revoluce s globálními důsledky pro mnoho oborů včetně tak specifických, jako je péče o duševní zdraví, a jeden z hlavních determinantů proměny mnoha vědních oborů a domén lidské společnosti. Mnohé výsledky aplikace umělé inteligence jsou příslibem pro budoucí zlepšení zdravotních systémů po celém světě, stejně jako pro zlepšení výsledků v oblasti péče o somatické a mentální zdraví významné části populace. Umělá inteligence může mít potenciál pro prosazování rovnosti v poskytování péče o psychické zdraví v tom smyslu, že jí učiní dostupnější. Bez ohledu na slibné výsledky v některých oblastech je na místě zdrženlivost a je nutné zvážit všechny etické důsledky jejího nasazení, aby se tak minimalizovalo riziko vzniku případných škod, a to zejména vůči zranitelnějším částem populace.

Cíl práce a výzkumná otázka

Umělá inteligence je obecně spojena s těmito základními otázkami:

- Jak umělá inteligence ovlivňuje lidi v každodenním životě a jak ho to změní?
- Jaká jsou možná rizika a důsledky umělé inteligence a jak se jim lze vyhnout?
- Jak by se měla veřejnost zodpovědně zabývat umělou inteligencí?
- Jaké etické problémy představuje vývoj a používání umělé inteligence?

Cílem práce je popsat a porovnat různé profesní etické normy, etické kodexy a různá etická vodítka pro vývoj a využití umělé inteligence a jejich prostřednictvím identifikovat možné etické problémy používání vybraných konverzačních nástrojů umělé inteligence používaných při psychoterapii pro zvládání úzkostných stavů.

Hlavní výzkumná otázka zní:

Jaké potenciální etické problémy představuje v současnosti vývoj a používání konverzačních chatbotů pro zvládání úzkostných stavů využívajících technologie umělé inteligence?

Vedlejší výzkumnou otázkou je:

Jaké jsou etické implikace předpokládaného budoucího masivnějšího používání konverzačních nástrojů užívajících technologie umělé inteligence v oblasti péče o duševní zdraví a před jakými případnými etickými výzvami a riziky stojí tvůrci systémů, zdravotníci a regulátoři zdravotní péče?

Aktuálnost tématu

Téma etiky užití umělé inteligence jako nástroje v psychoterapii je aktuální z několika závažných důvodů. V důsledku pandemie COVID-19, zhoršující se sociálně ekonomické situace, ale i strachu z válečného konfliktu v Evropě stoupá zejména u zranitelnější části populace komorbidita psychických poruch a onemocnění. Pandemie COVID-19 zesílila potřebu podpory péče o duševní zdraví napříč celým spektrem populace ve všech dotčených zemích. V některých zemích, kde globální požadavky přesahují kapacitu služeb podpory duševního zdraví a kde pro to byly vhodné technické a ekonomické podmínky, byly rozšířeny běžné psychoterapeutické postupy o inovativní prvky zvyšující kapacitu zdravotnických systémů. Příkladem může být inovace kognitivně behaviorální terapie při léčbě úzkostných stavů a poruch, v níž byly specifické části tradiční interakce mezi psychoterapeutem a pacientem nahrazeny nebo doplněny novými formami interakce využívající technologií umělé inteligence. Jednou z nezanedbatelných inovativních oblastí je vývoj konverzačních prostředků – chatbotů, využívajících specifickou technologii umělé inteligence pro vybrané procesy v psychoterapii a pro psychologické poradenství v oblasti zvládnání úzkostných a depresivních stavů a poruch. Již před pandemií vykazovaly tyto inovativní nástroje některé pozitivní výsledky. Vyvolávaly však také diskusi kvůli řadě etických a společenských výzev, které až do té doby nebyly významněji reflektovány, a to ani odbornou veřejností, ani regulátory zdravotních služeb či uživateli psychoterapeutické péče a psychologických konzultací. Téma mě zaujalo především tím, že je průnikem tří oblastí, na které jsem zaměřil svoje dosavadní vzdělávání i odborné působení, tj. oblast aplikace počítačových systémů, oblast psychologie a oblast aplikované etiky.

Použitá metoda

Pro teoretickou část práce jsem zvolil metodu deskriptivní analýzy a pro výzkumnou část exploračně komparační metodu spočívající ve výzkumu chování vybraných konverzačních nástrojů umělé inteligence v modelových situacích. Následuje identifikace potenciálních etických problémů z pohledu existujících etických kodexů, etických norem a obecně přijímaných etických principů, s nimiž může být chování konverzačních prostředků umělé inteligence v rozporu. Etika umělé inteligence užívaná v psychoterapii úzkostných stavů je dekomponována do bloků tvořících jednotlivé vrstvy této problematiky – umělá inteligence, psychoterapie, úzkostné stavy. Pro jednotlivé vrstvy jsou analyzovány výstupy z výzkumů publikovaných v impaktovaných periodikách. Společným jmenovatelem analýz propojujícím jednotlivé vrstvy je tematizace etiky.

Struktura práce

V první části je specifikováno základní pojmosloví, vymezeno pojmové a myšlenkové pole dané problematiky, v kterém se budu pohybovat. Druhá část obsahuje detailnější popis použité metody získávání informací a jejich analytického zpracování. Ve třetí části jsou srovnávána různá pojetí etiky v oblasti psychoterapie a umělé inteligence včetně různých metod řešení etických problémů a etických konfliktů. Čtvrtá část popisuje pojetí úzkostných stavů, jejich léčbu a hlavní faktory této léčby. Pátá část je věnovaná etické reflexi specifických konverzačních nástrojů umělé inteligence – chatbotů používaných v terapii. Šestá část se věnuje formulaci konkrétních etických důsledků a případných rizik při budoucím předpokládaném masovějším užití konverzačních prostředků umělé inteligence pro zvládnutí úzkostných stavů. Závěr je věnován shrnutí dané problematiky, sumarizaci výsledků a jejich limitů a nastínění možností dalšího zkoumání dané problematiky.

1. Základní pojmy a koncepty

1.1. Umělá inteligence

Umělá inteligence, jako zastřešující termín, se v různých kontextech používá různě a zahrnuje různé techniky a disciplíny. Definovat umělou inteligenci je obtížné také kvůli různorodým koncepcím lidské a mimolidské inteligence a rozdílnosti pojetí přirozenosti a umělosti. Obecně lze říci, že systémy umělé inteligence lze chápat jako technologie navržené k provádění činností, které obvykle vyžadují lidskou inteligenci¹. Termín umělá inteligence se používá i jako pojem pro oblast vědy, která zkoumá a rozvíjí koncepty specifické technologie. Umělou inteligenci lze tedy chápat jako „... vývoj, analýzu a simulaci výpočtově náročných, vysoce výkonných systémů pro provádění složitých úkolů, kde jsou úkoly široce definovány, zahrnují značnou flexibilitu a rozmanitost a jsou typicky podobné lidským aspektům poznání nebo vnímání“². Umělá inteligence je zároveň interdisciplinární přístup, který kombinuje biomedicínskou informatiku, matematiku, logiku, informatiku, inženýrství a další příbuzné obory³. Příkladnými odvětvími umělé inteligence jsou strojové učení, neuronové sítě, expertní systémy, rozpoznávání vzorů, vizuální analýza, zpracování jazyka a robotika. Pro přiléhavost k danému tématu a jednoduchost je zde užitá definice z dokumentu o strategii využití umělé inteligence, kde je umělá inteligence definována jako: „Technologie se schopností provádět úkoly, které by jinak vyžadovaly lidskou inteligenci, jako je vizuální vnímání, rozpoznávání řeči a překlad jazyka“⁴. Pro účely této práce je uvedená definice rozšířena v tom smyslu, že systémy umělé inteligence mají dnes obvykle schopnost učit se z nových podnětů a přizpůsobovat podle nich svoje algoritmy chování. Na umělou inteligenci lze pohlížet v

¹ DE SAINT LAURENT, Constance. In Defence of Machine Learning: Debunking the Myths of Artificial Intelligence. *Eur J Psychol* 2018. roč.14, č.4. s.734-747.

² BARDEN, John. Metaphor and Context: A Perspective from AI. (2009) [2022-05-31]. <<http://www.czso.czhttps://www.researchgate.net/publication/304736287>>.

³ PATEL, Vimla. The coming of age of artificial intelligence in medicine. *Artificial Intelligence in Medicine* 2009. č. 46, s. 5–17.

⁴ Department for Business, Energy and Industrial Strategy, Industrial Strategy: Building a Britain fit for the future (November 2017) [2022-05-31]. <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf> .

širším nebo užším smyslu. Obecná umělá inteligence je stroj s širokými kognitivními schopnostmi, schopný realizovat značný rozsah kognitivních procesů nebo alespoň přesvědčivě simulovat některé intelektuální schopnosti člověka a některé z nich díky své vysoké výpočetní výkonnosti i předčít. V užším smyslu plní systémy umělé inteligence vybrané úkoly, jaké by od člověka vyžadovaly užití jeho specifické části inteligence, jako je například rozpoznávání nesespecifických tvarů vybraných buněk. Tyto systémy jsou však omezené v rozsahu úkolů, které mohou vykonávat. V této práci, pokud to není uvedeno jinak, je použit pojem umělá inteligence v užším slova smyslu.

Termín umělá inteligence navrhl profesor matematiky John McCarthy, jenž v roce 1956 zorganizoval konferenci na téma, zda lze stroje naučit myslet⁵. Předložený návrh vycházel ze základní premisy výzkumu umělé inteligence, že pokud by se dala vlastnost inteligence, jako je učení, rozložit na jednotlivé části a operativně přesně definovat, pak by bylo možné vyrobit stroj, který by ji simuloval. Tato konference je známá jako jeden z počátečních milníků moderní éry umělé inteligence, včetně prvního setkání umělé inteligence, psychoterapie a psychologického poradenství, protože se zabývala kromě jiného i simulací netriviální konverzace. V mnoha ohledech se psychoterapeuti a psychologičtí poradci zabývají problémem účinné komunikace uskutečněné ve formě ústní, písemné, neverbální, výtvarné, ale také ve formě hudební v případě muzikoterapie. Deset let po konferenci byl vyvinut první program – textbot schopný komunikovat způsobem připomínajícím psychoterapeutický rozhovor. Tento debutový textbot Eliza z roku 1966, jehož název vychází ze slova robot, byl počítačový program, který vykonával automatizovanou konverzaci. Eliza byla navržena tak, aby simulovala rogeriánského terapeuta, a byla známá tím, že na otázky odpovídala otázkami, jak je to v terapii navržené Carl Rogersem za určité situace doporučováno⁶.

Pro lepší porozumění užití umělé inteligence v terapii, které bude dále v tomto textu popsáno detailněji, je také potřeba vymezit obsah pojmů strojové učení, neuronová síť a hluboké učení. Strojové učení je jedna z forem umělé inteligence, která dává aplikacím

⁵ MC CARTHY, John. A proposal for the Dartmouth summer research project on artificial intelligence. *AI Magazine* 2006, roč. 4, č 27, s. 12.

⁶ MAULDIN, Michael. ChatterBots, TinyMuds, and the Turing test: Entering the Loebner prize competition. (1994) [2022-05-31]. < <https://www.aaai.org/Papers/AAAI/1994/AAAI94-003.pdf> >.

umělé inteligence možnost učit se a zlepšovat se získávanými zkušenostmi, aniž by bylo nutné tyto aplikace modifikovat. Je-li k dispozici dostatečné množství dat, může se algoritmus strojového učení naučit předpovídat nebo řešit problémy, jako je například identifikace objektů na obrázcích nebo identifikace změn mimiky pacienta užívané v psychiatrii pro identifikaci suicidálního směřování pacienta nebo vyvodit změnu v jednání a postojích komunikujícího jedince, a na základě toho doporučit pacientovi specifickou změnu v jednání nebo i myšlení⁷. O etické problematičnosti takových doporučení bude pojednáno v páté kapitole.

Neuronová síť, jež vnesla podstatný impulz do kvalitativní změny umělé inteligence, je typ strojového učení volně inspirovaný organizací struktury nervových buněk neokortexu. Neuronová síť se skládá z jednoduchých procesních uzlů neboli umělých neuronů, které jsou vzájemně propojeny ve vrstvách. Každý uzel přijímá data z několika uzlů nad sebou a předává data dalším uzlům pod ním. Uzly přiřazují přijatým datům váhu a hodnotu. Pokud data neprojdou určitou prahovou hodnotou, nejsou předána jinému uzlu. Váhy a prahové hodnoty uzlů se upravují, když je algoritmus trénován, dokud podobný vstup dat nevede ke konzistentním výstupům⁸.

Hluboké učení je novější varianta neuronových sítí, využívající mnoho vrstev umělých neuronů k řešení složitějších problémů. Jeho akceptace odborníky na umělou inteligenci se díky výkonnosti výpočetní techniky od poloviny roku 2000 výrazně zvýšila, a proto dnes stojí za velkým zájmem o umělou inteligenci. Často se používá ke klasifikaci informací z obrázků, textu nebo zvuku včetně zvuků lidského hlasu⁹, což se ukazuje například jako významná funkčnost využívaná při detekci možné suicidality u komunikujícího pacienta napojeného na monitorovací systém.

⁷ An Encyclopædia Britannica Company MerriamWebster dictionary. [2022-05-31]. <<https://www.merriam-webster.com/dictionary/machine%20learning>> .

⁸ An Encyclopædia Britannica Company MerriamWebster dictionary. [2022-05-31]. <<https://www.merriam-webster.com/dictionary/neural%20networks>> .

⁹ IBM cloud. *AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference?* (27.5.2020) [2022-05-31]. <<https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>> .

Chatbot je software nebo počítačový program, který simuluje mezilidskou konverzaci prostřednictvím textových nebo hlasových interakcí¹⁰. Chatboti mají různé úrovně složitosti, jsou buď beze-stavové nebo stavové. Beze-stavoví chatboti přistupují ke každé konverzaci, jako by komunikovali s novým uživatelem. Naproti tomu stavoví chatboti mohou kontrolovat minulé interakce a vytvářet nové odpovědi v kontextu s předchozí konverzací. Největším kvalitativním posunem umělé inteligence byl přechod od expertních systémů založených na hierarchicky uspořádaných znalostech, které byly na své vrcholu v 80. letech minulého století, ke strojovému učení na počátku roku 2000.

1.2. Psychoterapie úzkostných poruch a úzkostných stavů

Úzkost je emoce charakterizovaná pocity strachu, neklidu, napětí, negativními myšlenkami a fyziologickými změnami, jako je zvýšený krevní tlak, bušení srdce, třes nebo pocení¹¹. Může to být obvyklá reakce organismu na blížící se vyšší duševní či fyzickou zátěž. Úzkost je možné pociťovat, když jedinec čelí obtížnému problému v práci nebo ve vztazích, před zkouškou nebo před důležitým rozhodnutím. Úzkost je emoce, která jedinci může pomoci, aby se účinně vyrovnal se zátěží tím, že mu pomůže dodat energii nebo mu pomůže soustředit se na přípravu k zátěži. Občasná úzkost je běžnou součástí života a po určité době vymizí. Pro lidi s úzkostnými poruchami však není úzkost dočasná a může podle stupně závažnosti poruchy vést až k ochromení v jednání, myšlení a prožívání. Příznaky poruchy mohou narušovat každodenní činnosti, jako je pracovní výkon, školní práce a vztahy. Existuje několik typů úzkostných poruch, včetně generalizované úzkostné poruchy, panické poruchy, sociální úzkostné poruchy a různých úzkostných poruch souvisejících s fobiemi.

Psychoterapie je léčebné působení psychologickými prostředky.¹² Terapeut působí prostřednictvím podnětů, které vytvářejí příznivé podmínky pro uzdravení. Indikace pro

¹⁰ An Encyclopædia Britannica Company *MerriamWebster dictionary*. [2022-05-31]. <<https://www.merriam-webster.com/dictionary/chatbot>>.

¹¹ PRAŠKO, Ján. *Úzkostné poruchy*. Praha: Portál, 2005.

¹² TIMULÁK, Ladislav. *Základy vedení psychoterapeutického rozhovoru*. Praha: Portál, 2005.

psychoterapii a výběr specifického postupu je součástí rozhodovacího procesu terapeuta, jehož výsledkem je přiřazení konkrétního pacienta s jeho vstupními zdravotními problémy, osobností, životní historií a sociálním okolím do určité vymezené psychoterapeutické péče, která je pro daného pacienta nejúčinnější. Na počátku je provedena obecná indikace, zda je vůbec možné pacienta léčit prostředky psychoterapie, a v další fázi specificko-prognostická indikace, v níž se rozhodne o přiřazení konkrétního psychoterapeutického postupu, formě terapie a konkrétního terapeuta. V průběhu terapie může dojít v závislosti na průběhu léčby též k adaptivní indikaci. Cílem je odstranění nebo zmínění obtíží nebo odstranění jejich příčin – obnovení zdraví, poznání a změna sebepoškozujícího jednání, účinné zvládnání problémů, konfliktů a životních úkolů. Má vést k uplatnění možností jedince, k naplnění životního smyslu. Cíle se různí v závislosti na druhu a charakteru terapie. Hlavními činnostmi a prostředky terapie je psychoterapeutický vztah, uspořádání situace v terapii, kdy by prostředí terapie mělo navozovat pocit bezpečí, rozhovor, při kterém je důležité fázování rozhovoru, kladení otázek, mlčení, udržování kontaktu, vyjadřování podpory, interpretace, konfrontace, shrnutí a reflexe, imaginace, učení. Vytvoření a podpora vhodného psychoterapeutického vztahu je jednou z problematických oblastí užití konverzačních prostředků umělé inteligence, tato problematika bude blíže analyzována v páté kapitole. Existují různé druhy terapie – dle různých hledisek direktivní a nedirektivní, symptomatická a kauzální, podpůrná a rekonstrukční nebo individuální, skupinová, hromadná. Podle příslušnosti k základním směrům rozlišujeme terapii psychodynamickou, behaviorální, humanistickou, kognitivní. Podle účinných faktorů a metod dělíme terapii na racionální, empatickou, abreaktivní, psychoanalytickou a interpersonální. Obecně lze říci, že u konverzačních prostředků umělé inteligence se používá terapie v jejím individuálním, direktivním a podpůrném nastavení a odpovídá podpůrné terapii v rámci kognitivně behaviorální terapie. V analytické části se proto zaměřím na rozbor kognitivně behaviorální terapie, z jejíž léčebného modelu vychází většina chatbotů¹³.

¹³ FULMER Russell. Artificial intelligence and counselling: Four levels of implementation. *Theory & Psychology* 2019, roč. 29, č. 6, s. 807–819.

1.3. Etika umělé inteligence v psychoterapii

Psychologie jako vědní obor a obor činností pomáhající profese se opírá o znalosti a dovednosti, které jsou výsledkem eticky vedeného výzkumu, a o etiku použití jeho výsledků v praxi, na niž se v kontextu řešené problematiky dále zaměřuji. Pro zajištění eticky přijatelného působení psychologa a psychoterapeuta je nutné usměrňování profesionálního chování prostřednictvím různých politik a regulativ¹⁴. Profesionální etika psychoterapeutů je formulována prostřednictvím zákonů, etických kodexů, předpisů a standardů, jimiž se daná profese řídí, a zastřešujících etických principů, které jsou základem vymahatelných standardů chování, dovedností etického rozhodování, strategií řízení rizik a seberegulace, včetně například emocionální kompetence. Pravidla a směrnice profesionálního chování jsou kodifikovány v etickém kodexu profese. Pro evropské psychology se jedná o Etický metakodex, vydaný Evropskou federací psychologických asociací (EFPA)¹⁵, zatímco pro severoamerické psychology jsou to Etické principy psychologů a kodex chování, které vydala Americká psychologická asociace (APA)¹⁶. Zákony a normy však nemohou řešit všechny problémy, s nimiž se profesionální psychologové a psychoterapeuti potýkají. Psychologové i psychoterapeuti se setkávají se situacemi, které jsou zcela jedinečné nebo závislé na kontextu, a proto je nemůže beze zbytku pokrývat relativně obecný etický princip, etická norma nebo etický kodex. Psychologové a psychoterapeuti se také mohou setkat se situacemi, které etický kodex nepředvídá, nebo se situacemi, kdy se zdá, že zastřešující etické principy kolidují se zákonem nebo politikou instituce, jež je zaměstnává. V důsledku toho se psychologové a psychoterapeuti musí spoléhat na překlenující etické hodnoty a při svém rozhodování z nich vycházet. Pomocí vhodného a systematického řízení rizik je možné snížit pravděpodobnost, že bude příslušný psycholog nebo psychoterapeut vyšetřován nebo odsouzen disciplinárním orgánem profesní komory. Strategie řízení rizik spojených s neetickým jednáním by měla být tedy zaměřena na implementaci nebo naplňování

¹⁴ LINDSAY, G. Etika pro evropské psychology. Praha: Triton, 2010. s.24.

¹⁵ European Federation of Professional Psychologist Association. *Meta-code of ethics*. Brussels: EFPA, 2005.

¹⁶ Ethical Principles of Psychologists and Code of Conduct. Washington DC: American Psychological Association (2020) [2022-05-31]. <<https://www.apa.org/about/policy/ethics>> .

zastřešujících etických principů¹⁷. To se týká i případů, kdy psychoterapeut nebo psycholog indikuje léčbu pomocí prostředku umělé inteligence.

¹⁷ Ethical Principles of Psychologists and Code of Conduct. Washington DC: American Psychological Association (2020) [2022-05-31]. <<https://www.apa.org/about/policy/ethics>>.

2. Popis užití metody

Pro teoretickou část práce byla zvolena metoda deskriptivní analýzy a pro výzkumnou část exploračně komparační metoda spočívající v pozorování chování vybraných konverzačních nástrojů umělé inteligence v modelových situacích a následná identifikace potenciálních etických problémů z pohledu etických kodexů, etických norem a obecně přijímaných etických principů. Problematika zkoumání umělé inteligence z perspektivy etičnosti jejího užití v psychoterapii úzkostných stavů byla rozložena do bloků tvořících jednotlivé vrstvy této problematiky. Pro tyto jednotlivé vrstvy byly analyzovány výstupy z výzkumů publikovaných v impaktovaných periodikách. V osmi elektronických databázích byla vyhledány recenzované odborné texty publikované před dubnem 2022 pomocí pojmů psychoterapie, etika a umělé inteligence, chatbot, úzkostné stavy, depresivní stavy a jejich souvisejících termínů. Data byla mapována do relačních struktur a byla provedena popisná a tematická analýza. Společným jmenovatelem analýz v jednotlivých vrstvách byla tematizace etiky. V prvním kroku byl na základě určeného kritéria proveden výběr studií řešících danou problematiku, ve druhém kroku se prováděla analýza a syntéza popisů daných témat ve vybraných studiích, ve třetím kroku byla provedena identifikace hlavních etických problémů souvisejících s technologiemi umělé inteligence v psychoterapii, ve čtvrtém kroku se realizovala analýza a komparace etických norem a profesních kodexů vztahujících se k problematice užití umělé inteligence v psychoterapii a psychologickém poradenství, pátý krok byl zaměřen na identifikaci etických problémů spojených s vybranými konverzačními aplikacemi umělé inteligence v psychoterapii úzkostných stavů, včetně gap analýzy vůči závěrům z předchozího kroku a v závěrečném kroku byly identifikovány etické implikace pro oblast budoucího vývoje a masovějšího užití umělé inteligence v psychoterapii úzkostných stavů. Konverzační aplikace umělé inteligence byly vybrána na základě hlavního selekčního kritéria, kterým bylo množství věrohodných testů účinnosti aplikací pro oblast zvládnutí úzkostných stavů vydaných v impaktovaných a recenzovaných publikacích.

3. Etika psychoterapie při využití umělé inteligence

3.1. Etika psychoterapie

V současnosti existují dva nejvlivnější etické kodexy pro oblast psychologie a psychoterapie, kterými jsou americký kodex společnosti American Psychological Association (APA) a evropský kodex federace European Federation of Psychologists' Associations (EFPA). Etický kodex APA¹⁸ poskytuje příslušnou vědomostní základnu pro odborníky pracující v oblasti psychologie, aby byli lépe vybaveni znalostmi, co dělat, když se setkají s nějakým typem etického dilematu. Některé z nich jsou principy nebo hodnoty, které by psychologové měli bezpodmínečně dodržovat. Etický kodex APA se skládá z pěti klíčových principů a deseti etických standardů. Principy představují vodítko, jež má pomoci inspirovat psychology při práci v jejich profesi, ať už pracují v oblasti péče o duševní zdraví, ve výzkumu nebo v podnikání v oblasti péče o psychické zdraví. Na druhé straně normy představují očekávané chování, které může při jejich porušení vést k profesionálním a právním důsledkům. Ne všechny etické problémy jsou jednoznačné, APA však nabízí psychologům hlavní principy, které by jim měly pomoci učinit v rámci jejich profese správná etická rozhodnutí.

Uvedené etické kodexy vycházejí vesměs z pěti¹⁹ bioetických principů formulovaných Beauchampem a Childressem – sebeurčení (autonomie), předcházení škodám (nonmaleficence), povinná péče (beneficence) a spravedlnost a specifická pozornost věnovaná vztahu mezi příjemcem a poskytovatelem léčby či péče²⁰. První princip etického kodexu APA *Beneficence a Nonmaleficence* stanovuje, že by se psychologové měli snažit chránit práva a blaho těch, s nimiž profesionálně pracují²¹. To zahrnuje klienty a pacienty, se kterými pracují v klinické praxi, ale i kdokoli jiný, s kým se zapojí do

¹⁸ Ethical Principles of Psychologists and Code of Conduct. Washington DC: American Psychological Association (2020) [2022-05-31]. <<https://www.apa.org/about/policy/ethics>> .

¹⁹ V české literatuře jsou obvykle uváděny čtyři principy nicméně poslední, osmé vydání Principles of Biomedical Ethics obsahuje principů pět. Pátý princip je zaměřen na charakter vztahu mezi příjemcem a poskytovatelem zdravotní péče.

²⁰ BEAUCHAMP, Tom L. – CHILDRESS, James F. Principles of Biomedical Ethics. Osmé, rozšířené vydání. New York: Oxford University Press, 2019. s. 99-370.

²¹ CONLIN, William. Ethical considerations for addressing distorted beliefs in psychotherapy. *Psychotherapy* 2019, roč. 56, č. 4, s. 449-458.

profesionální interakce. Tento princip převzatý z bioetiky²² nabádá psychology, aby se snažili eliminovat zaujatost, vztahy a předsudky, které by mohly negativně ovlivnit jejich práci. Zahrnuje to i nezávislé jednání ve výzkumu a zamezení vlivu sponzorů na výsledky výzkumu. Aplikace tohoto principu představuje určitý problém při použití konverzačních prostředků umělé inteligence, protože tyto prostředky zatím dostatečně nediferencují mezi etnickými, socio-kulturními různostmi pacientů a klientů a díky tomu může docházet ke krokům zhoršujícím účinnost terapie.

Druhý princip etického kodexu *Preciznost a odpovědnost* určuje, že psychologové mají morální odpovědnost za pomoc při zajišťování toho, že i ostatní pracovníci v jejich profesi dodržují příslušné etické standardy. Tento princip zdůrazňuje, že by práce psychologů měla zahrnovat i činnosti, které zlepšují etiku chování jejich kolegů. Tento princip je možné naplnit tím, že psycholog působí jako mentor, účastní se vzájemného hodnocení a upozorňuje na etická rizika nebo nesprávné chování kolegů. Pro naplňování tohoto principu musí psycholog a psychoterapeut, využívající konverzační prostředky umělé inteligence, provádět audit komunikace chatbotu s pacientem, a pokud identifikuje možná porušení tohoto principu, měl by informovat společnost zajišťující podporu chatbotu.

Třetí princip etického kodexu je *Integrita*. V praxi a výzkumu by se psychologové nikdy neměli pokoušet klamat nebo zkreslovat. Ve výzkumu může klamání zahrnovat vymyšlení nebo manipulaci s výsledky takovým způsobem, aby bylo dosaženo žádoucích výsledků. Psychologové by také měli ve své praxi usilovat o transparentnost a poctivost.

Čtvrtý princip etického kodexu je *Spravedlnost*, rovněž převzatý z bioetiky²³. Ten se v širším slova smyslu vztahuje k odpovědnosti být spravedlivý a nestranný. Tento princip říká, že všichni lidé mají právo na přístup a prospěch z pokroků, jichž bylo na poli psychologie dosaženo. Pro psychology je důležité, aby se k lidem chovali stejně a nedělali mezi nimi rozdíl například na základě socioekonomického postavení. Tento princip je

²² BEAUCHAMP, Tom L. – CHILDRESS, James F. *Principles of Biomedical Ethics*. Osmé, rozšířené vydání. New York: Oxford University Press, 2019, s. 155-256.

²³ BEAUCHAMP, Tom L. – CHILDRESS, James F. *Principles of Biomedical Ethics*. Osmé, rozšířené vydání. New York: Oxford University Press, 2019, s. 267-313.

jeden z nejdiskutovanějších ve spojení s konverzačními prostředky umělé inteligence, protože je to právě umělá inteligence, která by pomohla zpřístupnit vhodnou podpůrnou terapii i lidem, kteří nemají prostředky na úhradu standardní terapie nebo se jim terapie nedostává pro převis požadavků na terapii nad kapacitou psychoterapeutických služeb.

Pátý princip etického kodexu je *Respekt k lidským právům a důstojnosti*. Psychologové by měli respektovat právo na důstojnost, soukromí a důvěrnost těch, s nimiž profesně pracují.²⁴ Měli by se také snažit poznat a minimalizovat své vlastní předsudky a být si vědomi problémů vyplývajících z rozmanitosti a obav specifické komunity lidí v dané společnosti. Lidé mohou mít například specifické obavy související s jejich věkem, socioekonomickým postavením, rasou, pohlavím, náboženstvím, etnickým původem či zdravotním postižením. Aplikace tohoto principu může znamenat problém při použití konverzačních prostředků umělé inteligence, protože tyto prostředky nemusí dostatečně identifikovat etnickou, socio-kulturní a ekonomickou různost pacienta nebo klienta, a může proto v terapii docházet k ignorování těchto důležitých rozdílů.

Etický kodex²⁵ obsahuje rovněž deset standardů popisujících hlavní pravidla chování pro psychology pracující v klinické praxi anebo ve výzkumu. Tyto standardy bývají obecněji formulovány, aby pomohly vést chování psychologů v celé řadě oblastí a situací. Vztahují se na oblasti, jako je vzdělávání, terapie, reklama, soukromí, výzkum a publikace. První standard je *Řešení etických problémů*. Tento standard etického kodexu poskytuje informace o tom, co by psychologové měli dělat, aby vyřešili etické problémy, s nimiž se mohou při své práci setkat. To zahrnuje například rady, co by výzkumní pracovníci měli dělat, když je jejich práce zkreslována, a kdy nahlásit porušení pravidel etiky. Další standard je *Kompetence*. Je důležité, aby psychologové pracovali pouze v oblasti své odbornosti. Při léčbě klientů nebo při práci s veřejností musí psychologové mít na paměti, k čemu jsou vyškoleni, a to, k čemu vyškoleni nejsou, musí odmítnout. Tato norma stanoví, že v nouzové situaci mohou odborníci poskytovat služby, i když to

²⁴ SMITH, Randy. Psychotherapy in the #MeToo era: Ethical issues. *Psychotherapy* 2019, roč. 56, č. 7, s 483-490.

²⁵ Ethical Principles of Psychologists and Code of Conduct. Washington DC: American Psychological Association (2020) [2022-05-31]. <<https://www.apa.org/about/policy/ethics>> .

nespadá do rozsahu jejich praxe, aby byl zajištěn nezbytný přístup ke službám. Tento standard je důležitý i v kontextu konverzačních prostředků umělé inteligence, jejíž používání v klinické praxi musí předcházet proces kvalifikace psychoterapeutů pro její používání. Další standard představují *Lidské vztahy*. Ten je navržen tak, aby vedl psychology v jejich interakcích s ostatními pracovníky v oboru. To zahrnuje pokyny pro řešení sexuálního obtěžování, diskriminace a doporučuje postupy, jak se vyhnout nevhodným až vykořisťovatelským vztahům, jako je sexuální vztah s pacientem, pacientkou nebo podřízeným. Další standard *Soukromí a důvěrnost* popisuje povinnosti psychologů, pokud jde o zachování důvěrnosti informací týkajících se pacienta. Psychologové jsou povinni přijmout přiměřená opatření, aby informace o klientovi zůstaly díky jejich vhodné ochraně soukromé. Důvěrnost má však i svá omezení. Někdy psychologové potřebují sdílet informace o svých pacientech, například aby se mohli poradit s jinými odborníky na duševní zdraví. I když existují případy, kdy jsou informace předány neoprávněné osobě, musí se psychologové snažit tyto zásahy do soukromí a důvěrnosti minimalizovat. Tento standard má pro oblast konverzačních nástrojů umělé inteligence specifický aspekt navíc, kdy je nutné zavést takový technický systém ochrany dat, aby k těmto datům znemožnil přístup cizích osob. Navíc musejí být pravidelně auditovány přístupy k citlivým datům. Další standard *Reklama a další veřejná prohlášení* se týká psychologů, kteří inzerují své služby. Ten musí zajistit přesný popis jejich odbornosti, zkušenosti a školení. Psychologové se také musí vyhnout klamavým nebo nepravdivým marketingovým prohlášením. To platí i pro mediální obraz psychologů, poskytujících své odborné znalosti nebo prezentujících své názory v článcích, blozích, knihách či televizních pořadech. Další standard má název *Vedení záznamů a poplatky*. Vedení přesných záznamů je důležitou součástí práce psychologa, ať už pracuje ve výzkumu nebo s pacienty. Takový záznam má obvykle formu předepsanou stavovskou organizací, která zefektivňuje ad hoc audity. Standard *Záznamy o pacientech* zahrnují poznámky k případu a další diagnostická hodnocení používaná v průběhu léčby. Další standard je *Terapie*. Tento standard nastiňuje profesionální očekávání v kontextu poskytování terapie a zahrnuje důležitost získání informovaného souhlasu a vysvětlení léčebného procesu klientům. Řeší se důvěrnost a také některá omezení důvěrnosti, například když klient představuje pro sebe nebo ostatní bezprostřední nebezpečí. Pokud například psycholog musí z nějakého důvodu ukončit poskytování služeb klientovi,

očekává se, že takový psycholog svého klienta na změnu připraví a pomůže mu najít alternativní služby tak, aby nebyl narušen princip nonmaleficence.

Etický kodex obsahuje i popis vybraných etických zřetelů, kterých by měl psycholog dbát. Důvod je ten, že se psychologové často zabývají extrémně citlivými nebo nestabilními situacemi. První etický zřetel se týká *prospěchu klientů*. Psychologové často pracují s jednotlivci, kteří jsou zranitelní kvůli svému věku, postižení, intelektuálním schopnostem a dalším obavám. Při práci s těmito jedinci se psychologové musí vždy snažit chránit prospěch svých klientů. Další etický zřetel se týká *informovaného souhlasu*. Když jsou lidé v roli příjemců psychologických služeb, mají právo vědět, co mohou očekávat. V terapii proces získávání informovaného souhlasu zahrnuje nejen vysvětlení, jaké služby jsou nabízeny a jaká mohou být možná rizika, ale i právo pacienta opustit léčbu. Dalším etickým zřetelem je *důvěrnost*. Terapie vyžaduje poskytnutí bezpečného místa pro klienty, kde mohou diskutovat o hluboce osobních otázkách, aniž by se museli bát, že budou tyto informace sdíleny s ostatními nebo budou zveřejněny. Někdy však může nastat situace, kdy psycholog potřebuje sdílet některé podrobnosti, například když konzultuje s jinými odborníky nebo když publikuje svůj výzkum. Pro tyto případy existují *Etické pokyny*, které určují, kdy a jak mohou být některé informace sdíleny, a také stanovují některé kroky, které by psychologové měli k ochraně soukromí svých klientů podniknout. Posledním zřetelem je *Kompetence*. Důležitým etickým zájmem jsou také školení, vzdělávání a zkušenosti psychologů. Ti musí mít dovednosti a znalosti, aby mohli správně poskytovat služby, které klienti potřebují.

3.2. Etika umělé inteligence v perspektivě bioetiky

To, že bude mít umělá inteligence zásadní dopad na společnost, už není v současné době nijak zpochybňováno. Současná debata se proto místo toho soustřeďuje na to, do jaké míry bude tento dopad pozitivní nebo negativní, pro koho, jakými způsoby, na jakých místech a v jakém časovém horizontu. Vývoj umělé inteligence a její použití otvírá

společnosti čtyři hlavní příležitosti²⁶. Tyto příležitosti se významně dotýkají lidské důstojnosti a rozvoje jedince i společnosti v tom, že budou mít vliv na to, kým se můžeme stát, co můžeme udělat, čeho můžeme dosáhnout a jak můžeme rozvíjet vztahy mezi námi a světem. Umělou inteligenci lze využít k podpoře lidské důstojnosti a rozvoji možností člověka, může ale dojít i k jejímu nadužívání nebo zneužívání. Tím může docházet ke vzniku rizik. Na druhou stranu však strach, nevědomost, neopodstatněné obavy nebo přehnaná reakce mohou vést společnost k tomu, že technologie umělé inteligence nevyužije svůj plný potenciál, což může způsobit významné náklady ze ztráty příležitosti. Může to zahrnovat například těžkopádnou nebo nesprávně koncipovanou regulaci, nedostatečné investice nebo neopodstatněnou veřejnou reakci. Je ovšem nutné také zvážit rizika spojená s neúmyslným nadužíváním nebo úmyslným zneužíváním technologií umělé inteligence, založeným například na nesprávných pobídkách, nezodpovědné ziskuchtivosti společností vyvíjejících a produkujících příslušné systémy, nepřátelské geopolitice různých států nebo nekalých úmyslech. Některá tato rizika se v současné době již zčásti naplňují. Možnost společenského pokroku, již představují výše uvedené příležitosti, musí proto být zvážena spolu s rizikem, že umělá inteligence umožní nebo posílí neeticky motivované jednání. Reálným rizikem je však i to, že umělá inteligence může být nedostatečně využívána ze strachu z nadměrného nebo nesprávného použití.

K přijetí a osvojení technologií umělé inteligence by mělo nastat pouze tehdy, budou-li přínosy eticky přijatelné, smysluplné a potenciální rizika preventabilní a minimalizovatelná prostřednictvím řízení rizik nebo garancí nápravy při jejich naplnění. Tyto postoje budou záviset na zapojení širší společnosti do úvah o směřování a využívání umělé inteligence a na otevřenosti ve vztahu k tomu, jak funguje, a také na srozumitelných, snadno dostupných informacích o mechanismech regulace a nápravy při naplnění rizika. Tímto způsobem lze etický přístup k umělé inteligenci vnímat i jako systém včasného varování před riziky, která mohou ohrozit jedince i společnost. Jedná se tedy o nový přístup k aplikované etice, která mnohdy sledovala a následovala vývoj

²⁶ FLORIDI, Luciano. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds & Machines* 2018, roč. 28, s. 689–707.

poznání v jednotlivých oborech a která by se měla více zaměřovat na budoucnost a budoucí vývoj tak, jak to doporučuje autor prospektivní bioetiky Josef Kuře²⁷.

3.3. Etika umělé inteligence v psychoterapii

Užití umělé inteligence v psychoterapii a psychologickém poradenství má etické dopady a vyvolává řadu ontologických a epistemologických otázek. Jedna z podmínek pro to, aby mohla umělá inteligence substituovat některé činnosti psychoterapeuta nebo psychologického poradce, je, že musí mít schopnost se učit. Stroje, které se učí, jsou paradigmaticky odlišné od standardních strojů. Hlavní odlišnost je v kontrole složek ovládajících stroj. Člověk, jenž konstruuje a využívá standardní stroj, si nad takovým strojem zachovává kontrolu. K nehodám dochází u strojů například při automobilové nehodě, ale ani tehdy nemusí nehodu nutně způsobit řidič vozidla. Na vině může být lidská chyba v konstrukci, při výrobě vozidla či v datech navigace nebo jen nepříznivé počasí. K nehodám nedochází proto, že by automobil udělal špatné rozhodnutí. Naopak stroj, který se učí prostřednictvím svých vlastních zkušeností, může mít dovednosti a schopnosti neznámé jeho lidským původcům. Jedním z příkladů je AlphaGo, počítačový program určený ke hraní deskové hry Go²⁸. Hraním tisíců her proti lidským konkurentům a ostatním počítačům se AlphaGo zlepšil do té míry, že v roce 2016 porazil mistra světa ve hře Go. Vítězství AlphaGo nad člověkem je považováno za milník v historii strojového učení, protože Go je známá jako hra vyžadující nejen memorování nazpaměť, ale také strategii a intuici. AlphaGo si osvojil myšlenkovou autonomii a navíc prokázal, že zařízení s umělou inteligencí využívající strojové učení mohou z pohledu člověka jednat nepředvídatelným způsobem, a dokonce překonávat lidi v taktické zdatnosti, jako je například vyhodnocení různých potenciálních scénářů. Strojové učení je z hlediska svých možností nejspíš teprve v počátcích, i tak už ale vyvolává řadu otázek pro psychoterapeutickou a poradenskou profesi. Pokud se například mohou psychoterapeuti

²⁷ KUŘE, Josef. Prospektivní bioetika, Studie ke konceptu etiky biomedicínských technologií. Praha: Filosofie 2020.

²⁸ GIBNEY, Elisabeth. Google AI algorithm masters ancient game of Go. *Nature News* 2016, č.529, s. 445–446.

v rámci psychoterapeutického výcviku učit, zlepšovat své dovednosti a nakonec poskytovat psychoterapeutické intervence, může umělá inteligence dosáhnout učením stejného souboru dovedností, a navíc se oproti psychoterapeutům učí významně rychleji. Jak tento fakt autonomnosti umělé inteligence může ovlivnit obor psychoterapie? Stejně jako Go i psychoterapie a poradenství zahrnují intuici a strategii. Mohla by pokročilá umělá inteligence zčásti fungující jako psychoterapeut a poradce zpochybnit terapeutický postup i zkušeným psychoterapeutům?

Zatímco některé ontologické otázky spojené s umělou inteligencí nemusí být pro psychoterapeuty zcela zásadní, otázka navržená otcem zakladatelem umělé inteligence Alanem Turingem je zjevně relevantní. Turing navrhl výzkumnou otázku: „*Jak dobře může stroj napodobovat lidskou konverzaci?*“²⁹. Turingův test staví počítačový systém proti lidské subjektivní zkušenosti. Tento test, známý jako The Imitation Game, instruuje lidské účastníky testu, aby prostřednictvím textu interagovali s neznámou entitou. Entitou může být počítačový program nebo lidská bytost, která komunikuje s účastníkem písemnou formou. Pokud účastník uhodne, že konverzuje s počítačem, tak počítačový program simulující lidskou reakci testem neprojde. A naopak, pokud počítač dostatečně napodobuje lidskou konverzaci a účastníka, tak program testem projde. V oblasti, kde se intenzivně investuje do lidské konverzace, se Turingův test může ukázat jako klíčový při zvažování implementace umělé inteligence do poskytování vybrané části psychoterapie a poradenství. Tímto postupem by psychoterapeuti a poradci mohli přispět k rozhodování, že používání psychologické umělé inteligence jako doplňku k tradiční psychoterapii a poradenství může být přínosem jak pro klienty, tak i pro profesi psychoterapie a poradenství.

Pro uvažování o použití aplikací umělé inteligence v oblasti péče o psychické zdraví jsou podstatná relevantní etická vodítka a etické principy pro identifikaci etických problémů při užití aplikace umělé inteligence při terapeutické intervenci³⁰. Umělá inteligence patří mezi nejrychleji se rozvíjející oblasti pokročilých technologií v medicíně

²⁹ OPPY, Graham - DOWE David , "The Turing Test", The Stanford Encyclopedia of Philosophy (2021) [2022-05-31] . <<https://plato.stanford.edu/archives/win2021/entries/turing-test/>>.

³⁰ BURUK, Banu. A critical perspective on guidelines for responsible and trustworthy artificial intelligence. *Med Health Care and Philos* 2020. roč. 4 čís. 23, s. 387–399.

a v psychoterapii. Nejdůležitější vlastností umělé inteligence, která ji odlišuje od ostatních produktů pokročilých technologií, je její schopnost optimalizovat svůj původní program a rozhodovací algoritmy prostřednictvím schopností hlubokého učení na základě nově získaných dat. Tento rozdíl v dynamice je důvodem, proč technologie umělé inteligence zřetelně vystupuje z etických problémů jiných pokročilých technologických produktů. Etické otázky technologie umělé inteligence se pohybují od ochrany soukromí a důvěrnosti osobních údajů až po etický status a hodnotu prvků umělé inteligence v závislosti na jejich schopnosti hlubokého učení a rozsahu oblastí, ve kterých působí. Základem pro naplňování těchto nároků byl rozvoj etických norem a vodítek pro plánování, vývoj, produkci a používání technologií umělé inteligence.

V tomto ohledu bylo vytvořeno několik dokumentů zaměřených na deklaraci hodnot a doporučení v dalším rozvoji umělé inteligence. Doporučení Rady pro umělou inteligenci při Organizaci pro hospodářskou spolupráci a rozvoj (OECD)³¹ je jeden z mnoha dokumentů zveřejněných v posledních letech, které obsahují doporučení ohledně zásad etiky umělé inteligence. Tyto dokumenty byly vytvořeny také podnikovou sférou (Google, IBM, Microsoft), nebo institucemi vlád (Montrealská deklarace, skupina expertů Evropské komise) a existuje i dokument z akademické obce AI4People³². S ohledem na příslušnost Česka k Evropské unii je další výklad spojen zejména s doporučeními skupiny expertů Evropské unie.

Evropská komise jmenovala skupinu expertů, kteří mají poskytovat rady ohledně její strategie umělé inteligence. Během prvního roku svého mandátu pracovala skupina expertů pro umělou inteligenci (AI HLEG) na dvou hlavních výstupech. První dokument *Etické pokyny pro důvěryhodnou AI*³³ předkládá přístup k umělé inteligenci zaměřený na člověka a klíčový seznam sedmi požadavků, které by systémy umělé inteligence měly splňovat, aby byly důvěryhodné.

³¹ Recommendation of the Council on Artificial Intelligence. (2019) [2022-5-31]. <<https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>>.

³² FLORIDI, Luciano. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds & Machines* 2018, roč. 28, s. 689–707.

³³ Ethics guidelines for trustworthy AI. (8.4.2019) [2022-05-31]. <<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>>.

Důvěryhodnost umělé inteligence je založena na třech požadavcích. Podle nich by měl být produkt užívající umělou inteligenci legální, tj. respektovat veškeré platné právní a správní předpisy, měl by být etický, tj. zajišťovat dodržování etických zásad, a měl by být robustní, a to z technického i sociálního hlediska. V souladu s přístupem založeným na základních právech jsou určeny etické zásady a související hodnoty, které musí být dodržovány při vývoji, zavádění a používání systémů umělé inteligence. Mezi základní pokyny patří požadavek: „...vyvíjet, zavádět a používat systémy umělé inteligence způsobem, jenž dodržuje etické zásady týkající se: respektování lidské autonomie, spravedlnosti a vysvětlitelnosti a předcházení újmám. Uznat potenciální konflikty mezi těmito zásadami a zabývat se jimi“³⁴. Druhý pokyn zdůrazňuje, že by se měla věnovat zvláštní pozornost situacím, v nichž vystupují zranitelnější skupiny obyvatel, jako jsou děti, osoby se zdravotním postižením a jiné osoby, které byly v minulosti znevýhodněny nebo jsou ohroženy vyloučením, a situacím, v nichž existuje mocenská nebo informační asymetrie, jako například ve vztazích mezi zaměstnavateli a zaměstnanci nebo mezi podniky a spotřebiteli. Poslední základní pokyn doporučuje uvědomit si a zohlednit skutečnost, že přes veškeré výhody, které jednotlivcům i společnosti obecně přinášejí, představují systémy umělé inteligence také určitá rizika a mohou mít negativní dopad, včetně dopadů, které lze obtížně předjímat, určit nebo měřit, jako je například dopad na demokracii, právní stát, spravedlivé rozdělování nebo na samotnou lidskou mysl.

Evropská směrnice obsahuje i pokyny k realizaci důvěryhodné umělé inteligence a uvádí sedm požadavků, které by měly systémy umělé inteligence splňovat a které mají ve svých částech i etický přesah³⁵. Tento požadavek klade na designera a provozovatele nárok zajistit, aby byly při vývoji, zavádění a používání systémů umělé inteligence splněny požadavky na důvěryhodnou umělou inteligenci v těchto oblastech:

- 1) lidský faktor a dohled – zahrnuje lidská práva, lidský faktor a lidský dohled

³⁴ Tamtéž s.2.

³⁵ Tamtéž s.3

- 2) technická robustnost a bezpečnost – zahrnuje odolnost vůči útokům a zabezpečení, nouzový plán a obecnou bezpečnost, přesnost, spolehlivost a reprodukovatelnost
- 3) ochrana soukromí a správa dat – zahrnuje ochranu soukromí, kvalitu a integritu údajů a přístup k údajům
- 4) transparentnost – zahrnuje sledovatelnost, vysvětlitelnost a komunikaci
- 5) rozmanitost, nediskriminace a spravedlnost – zahrnuje předcházení nespravedlivé podjatosti tvůrců systémů umělé inteligence, přístupnost a univerzální design a zapojení zúčastněných stran
- 6) dobré environmentální a sociální podmínky – zahrnuje udržitelnost a šetrnost k životnímu prostředí, sociální dopad, společnost a demokracii
- 7) odpovědnost – zahrnuje auditovatelnost případné odpovědnosti za škodu, minimalizaci negativních dopadů a podávání zpráv o těchto dopadech a možnost zjednání nápravy.

Druhý dokument vypracovaný skupinou expertů *Zásady a investiční doporučení pro důvěryhodnou AI*³⁶ je zaměřený na realizaci projektů umělé inteligence a na jejich financování. Výsledky práce expertní skupiny byly prezentovány na prvním evropském shromáždění umělé inteligence v červnu 2019. Po tomto shromáždění prodloužila Evropská komise mandát skupiny o další rok. Tento rozšířený mandát umožnil skupině pokračovat v práci a zahájit pilotní provoz pravidel obsažených v dokumentu *Etické pokyny pro důvěryhodnou AI*³⁷. Mandát AI HLEG skončil v červenci 2020 představením dalších dvou výstupů: *Konečný seznam hodnocení důvěryhodné AI a Odvětvové úvahy o politice a investiční doporučení*. Dokument zkoumá možnou implementaci doporučení, která skupina dříve zveřejnila, ve třech konkrétních oblastech použití: Veřejný sektor, Zdravotní péče a Internet věcí.

³⁶ Policy and investment recommendations for trustworthy Artificial Intelligence. (26.6.2019) [2022-05-31]. < <https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence> >.

³⁷ Communication: Building Trust in Human Centric Artificial Intelligence. (8.4.2019) [2022-05-31].< <https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence> >.

Další normu pro etický vývoj a provoz systémů využívajících umělou inteligenci, která je zčásti komplementární ke směrnici Evropské unie, navrhli specialisté společnosti Google³⁸. Norma je koncipována podle principu primární orientace na cíl a stanovuje určující cíle umělé inteligence včetně vodítek, jak jich dosáhnout. Norma vychází z toho, že umělá inteligence by měla být především společensky prospěšná, protože dosah nových technologií se stále více dotýká společnosti jako celku. Vodítkem je to, že pokroky v umělé inteligenci budou mít transformační dopady v celé řadě oblastí, včetně zdravotnictví, bezpečnosti, energetiky, dopravy, výroby a zábavy. Při zvažování potenciálního rozvoje a využití technologií umělé inteligence je proto nutné vzít v úvahu širokou škálu sociálních a ekonomických faktorů a je nutné postupovat tam, kde se lze domnívat, že celkové pravděpodobné přínosy podstatně převýší předvídatelná rizika a nevýhody. Norma také zdůrazňuje to, že umělá inteligence mimo jiné zlepšuje schopnost společnosti porozumět významu obsahu ve velkém měřítku, resp. ve velkém objemu dat. Je potřeba usilovat o to, aby pomocí umělé inteligence byly snadno dostupné vysoce kvalitní a přesné informace, přičemž je nutné respektovat kulturní, společenské a právní normy v zemích, kde se umělá inteligence používá. Google se v této normě zavazuje, že bude promyšleně vyhodnocovat, kdy zpřístupnit příslušné technologie na nekomerční bázi. Mezi cíle s explicitně etickým charakterem patří zamezení takovému vývoji nástrojů umělé inteligence, které by vytvářely nebo posilovaly nežádoucí předsudky. Algoritmy a datové sady umělé inteligence mohou odrážet, posilovat nebo i omezovat nespravedlivé předsudky. Odlišení spravedlivých předsudků od nespravedlivých není vždy jednoduché a liší se v různých kulturách a společnostech. Proto je nutné se snažit zabránit nespravedlivým dopadům na lidi, zejména těm, které se týkají citlivých charakteristik, jako je rasa, etnická příslušnost, pohlaví, národnost, příjem, sexuální orientace, schopnosti a politické nebo náboženské přesvědčení. Třetí cíl obsahující rovněž etickou dimenzi je zaměřen na bezpečnost a testování z hlediska bezpečnosti. Vodítkem je nutnost pokračovat ve vývoji a uplatňování přísných bezpečnostních postupů, aby se zamezilo dosažení nezamýšlených výsledků, které vytvářejí riziko poškození uživatele umělé inteligence. Systémy umělé inteligence musejí být navrhovány tak, aby byly

³⁸ Responsible AI practices. [2022-05-31]. < <https://ai.google/responsibilities/responsible-ai-practices/> >.

náležitě odolné proti nahodilostem, a musejí být vyvíjeny v souladu s osvědčenými postupy ve výzkumu bezpečnosti umělé inteligence. Další cíl je spojen s odpovědností vůči lidem využívajícím umělou inteligenci. Vodítkem je požadavek na takové systémy umělé inteligence, které poskytují vhodné příležitosti pro zpětnou vazbu, relevantní vysvětlení a jsou uživatelsky přátelské. Pátým cílem je začlenit do vývoje umělé inteligence zásady návrhu ochrany soukromí. Vodítkem je volba uživatele pro upozornění a informovaný souhlas včetně transparentnosti a kontroly nad používáním dat. Posledním cílem je dát k dispozici technologie společnosti Google, které jsou v souladu s uvedenými zásadami, a umožnit tak mnohostranné využití s tím, že je nutné pracovat na omezení potenciálně škodlivých nebo zneužívajících aplikací. Společnost Google rovněž definuje zásady, podle nichž se bude rozhodovat při eticky sporných nasazení umělé inteligence. Hlavní zásadou vycházející z principu nonmaleficence je, že nebude technologie umělé inteligence nasazována v technologiích, které mohou způsobit celkovou škodu, resp. tam, kde existuje významné riziko poškození. Nasazení není možné tam, kde se umělá inteligence používá pro zbraňové systémy nebo jiné technologie, jejichž hlavním účelem je způsobit zranění osob. Totéž platí pro technologie, které shromažďují nebo využívají informace ke sledování, porušující mezinárodně uznávané normy, a technologie, jejichž účel je v rozporu se široce uznávanými zásadami mezinárodního práva a lidských práv.

I přes určité etické konotace je patrné, že z etické perspektivy jsou tyto cíle a vodítka zaměřeny spíše obecně a technicistně. Některá z nich mají spíše rétoricky deklarativní charakter, který se snaží marketingovým žargonem vytvořit pozitivní obraz o společnosti Google a který bez dalších zpřesnění nemůže účelně pomoci řešit etickou problematiku v konkrétních oborech aplikované etiky, jako je bioetika nebo environmentální etika.

4. Koncept terapie úzkostných poruch a stavů

4.1. Klinický obraz úzkosti

Úzkost je obvyklá reakce lidské psychiky na fyzickou nebo psychickou zátěž, a pokud se pohybuje v mezích normality, je v běžných situacích prospěšná. Může nás upozornit na nutnost soustředit pozornost na vhodnou přípravu, jak se vyrovnat se zátěží nebo jak se vyhnout nebezpečí. Úzkostné poruchy se liší od normálních pocitů nervozity nebo úzkosti a zahrnují nadměrný strach nebo nadměrnou a déletrvajících úzkost. Je pro ně charakteristické, že je obtížně stanovitelný její spouštěč. Úzkostné poruchy jsou nejčastějšími duševními poruchami a v určité fázi života postihují až 24.9 procent dospělých³⁹. Úzkostné poruchy jsou léčitelné a k dispozici je řada účinných léčebných postupů, přičemž léčba pomáhá většině lidí vést normální produktivní život. Úzkost nás může upozornit na nebezpečí a pomoci nám připravit se a věnovat pozornost případné reakci na ohrožení. Úzkostné poruchy mohou způsobit, že se lidé budou snažit vyhýbat situacím, které spouštějí úzkosti nebo zhoršují jejich příznaky. Ovlivněn může být pracovní výkon, školní práce a osobní vztahy. Síla prožitku úzkosti i strachu může být různá. Může být pocíťována jen jako slabá nervozita nebo nepohoda, nebo v extrémním případě jako zahlcující panika a hrůza. Průběh úzkosti může být kontinuální s průměrnou mírně kolísající intenzitou anebo se projevuje v záchvatech. Spontánní záchvat paniky nemusí mít zjevnou příčinu nebo spouštěč a může se objevit náhle. Naproti tomu, pokud existuje zjevná příčina, nehovoříme o panice, ale o fobii nebo strachu⁴⁰. Úzkost zažívá čas od času každý. Může to být důsledek sledování podezřelého pohybu nebo sledování horroru v televizi, důsledek prožitku nebo svědectví trestného činu. Každý může mít v těchto situacích úzkost, ale osoba s úzkostnou poruchou zažívá přetrvávající nebo opakující se úzkost, která mu brání v plné účasti na životě. Úzkost se může pohybovat od relativně mírné emoce, kdy je prožívána nervozita doprovázená pocitem neklidu, až po těžké emoce, kdy se jedná o časté a zneschopňující záchvaty paniky. Těžké úzkostné poruchy mohou vést k tomu, že člověk změní svůj životní styl, a aby úzkosti vyhověl,

³⁹ PRAŠKO, Ján. *Úzkostné poruchy*. Praha: Portál, 2005. s. 9.

⁴⁰ OCISKOVÁ Marie., PRAŠKO, Ján. *Generalizovaná úzkostná porucha*. Praha: Grada, 2017. s. 25.

není například schopen odejít z domova. Detailnější rozlišení Beckova modelu popisuje strach jako neurofyziologický stav poplachu zahrnující kognitivní ohodnocení hrozícího stresu nebo ohrožení bezpečí jedince. Naproti tomu úzkost je komplexní kognitivní, afektivní, fyziologická a behaviorální reakce aktivovaná tehdy, kdy očekávané události nebo okolnosti jsou považovány za vysoce averzivní, protože jsou vnímány jako nepředvídatelné, nekontrolovatelné události, které by mohly potenciálně ohrozit životně důležité zájmy jednotlivce⁴¹. Pro diagnostiku poruchy podle MKN-10 je nutná přítomnost úzkostných příznaků minimálně po dobu 6 měsíců⁴².

4.2. Prevalence úzkostné poruchy

Úzkostné poruchy jsou po závislosti na návykových látkách nejčastějšími psychiatrickými poruchami a jsou spojeny s vysokou zátěží nemocí. Velká část pacientů zůstává nediodagnostikovaná, neléčená nebo je léčena neadekvátně. Rovněž bývá častá komorbidita s jinými úzkostnými nebo afektivními poruchami. Věk nástupu úzkostných poruch se u jednotlivých poruch liší. Úzkostné poruchy mají tendenci mít chronický průběh, přičemž symptomy kolísají v závažnosti mezi obdobími relapsu a remise. Po 50. roce věku byl v epidemiologických studiích pozorován výrazný pokles prevalence úzkostných poruch. Jediná úzkostná porucha, která je stále běžná u lidí ve věku 50 a více let, je generalizovaná úzkostná porucha⁴³.

4.3. Etiologie úzkostných stavů a úzkostných poruch

Současná konceptualizace etiologie úzkostných poruch zahrnuje interakci psychosociálních faktorů, např. nepřízeň dítěte u rodičů, stres nebo trauma či genetickou zátěž a zranitelnost, která se projevuje neurobiologickými a neuropsychologickými

⁴¹ CLARK, David A.- BECK, Aaron T. *Cognitive therapy of anxiety disorders: Science and practice*. New York: Guilford Press, 2011.

⁴² International Statistical Classification of Diseases and Related Health Problems, Geneva: WHO, 2016.

⁴³ OCISKOVÁ Marie., PRAŠKO, Ján. *Generalizovaná úzkostná porucha*. Praha: Grada, 2017. s. 22-23.

dysfunkcemi. Etiologie úzkostných a afektivních poruch byla tradičně zaměřena na genetické faktory. V poslední době se hromadí důkazy naznačující, že epigenetické faktory, nezávisle na variacích sekvence DNA, určují průběh afektivních poruch, panické poruchy, posttraumatické stresové poruchy, sociální úzkostné poruchy a také úzkosti z vysazení po drogové závislosti. Drogy, chronický stres i traumatické vzpomínky kódují trvalé epigenetické přeprogramování, které mění interakci mozku s prostředím a vede k dysfunkčnímu chování. Vystavení látce nebo podmínkám prostředí vyvolávajícím úzkost a afektivní symptomy může vyvolat epigenetickou dysregulaci synaptické plasticity prostřednictvím neurotrofických faktorů⁴⁴.

4.4. Léčba úzkostných stavů a úzkostných poruch

Úzkostné poruchy se léčí psychoterapií, psychofarmaky nebo oběma postupy současně⁴⁵. Existuje několik psychoterapeutických léčebných postupů pro léčbu úzkosti, přičemž výběr vhodného postupu by měl být proveden ve spolupráci s psychologem nebo psychiatrem. Základní postup psychoterapie, která může pomoci lidem s méně závažnou úrovní úzkostné poruchy, je terapie rozhovorem. Aby byla psychoterapie účinná, musí být zaměřena na specifické úzkosti pacienta a musí být přizpůsobena jeho potřebám. Jedním typem psychoterapie, která může pomoci lidem s úzkostnými poruchami a která tvoří i východisko terapeutické metody užívané v designu aplikací umělé inteligence pro zvládnání úzkostných stavů, je například kognitivně-behaviorální terapie. Ta učí lidi různým způsobům myšlení, chování a reakcím na vybrané situace, aby jim pomohla cítit se v podobných situacích méně úzkostlivě a jednat racionálněji. Základní metoda kognitivně-behaviorální terapie, používaná k léčbě úzkostných poruch, je expoziční terapie. Ta se zaměřuje na konfrontaci s obavami, které jsou základem úzkostné poruchy, aby pomohla lidem zapojit se do činností, jimž se vyhýbali. Expoziční terapie se někdy používá spolu s relaxačními cvičeními. Takovou aplikaci představují i některé

⁴⁴ RÈGUE-GUYON, Mathilde. Neuroepigenetics of Neurotrophin Signaling: Neurobiology of Anxiety and Affective Disorders, *Progress in Molecular Biology and Translational Science*, 2018, s. 159-193.

⁴⁵ PRAŠKO, Ján. *Úzkostné poruchy*. Praha: Portál, 2005.

konverzační nástroje umělé inteligence. Terapie pomocí přijetí a závazku, případně terapie pomocí všímavosti jsou další možností léčby některých typů úzkostných poruch⁴⁶. Terapie pomocí přijetí a závazku a všímavosti, která je součástí třetí vlny kognitivně-behaviorální terapie, má k negativním myšlenkám indukovaným úzkostí jiný přístup než klasická kognitivně-behaviorální terapie. Ke snížení neklidu a úzkosti používá strategie, jako je přijetí negativních myšlenek, stanovení cílů a aplikaci všímavosti vůči prožívaným mentálním obsahům.

4.5. Metody léčby úzkostných stavů a úzkostných poruch

Základní metoda léčby v kognitivně behaviorální terapii je expozice. Techniky založené na expozici jsou některé z nejběžněji používaných metod kognitivně behaviorální terapie používaných při léčbě úzkostných poruch. Jeden teoretický rámec pro pochopení zdůvodnění léčby založené na expozici vychází z teorie emočního zpracování.⁴⁷ Podle této teorie je strach reprezentován asociativními sítěmi, které jsou kognitivními strukturami strachu, udržujícími informace o obávaném podnětu, reakcích na strach, jako je například vyhýbání se, nebo psychofyzilogické reakce, jako je pocení, zvýšená srdeční frekvence a význam podnětů a automatických reakcí, kdy je například zvýšená srdeční frekvence považována automaticky za infarkt. Když se pacient v prostředí setká s podnětem, který se podobá obávanému podnětu, aktivují tyto asociativní síť strachu. Struktura strachu je patologická, když vztah mezi podněty, reakcemi a jejich významem neodpovídají realitě, například když je aktivován pro bezpečné podněty nebo reakce, které se podobají těm obávaným. Kromě toho je struktura strachu udržována vyhýbavým chováním, které neumožňuje nové učení. Expozice je navržena tak, aby modifikovala strukturu patologického strachu tím, že ji nejprve aktivuje a poté poskytne nové informace, které vyvrací patologické, nerealistické

⁴⁶ ŠLEPECKÝ, Miloš - PRAŠKO, Ján – KOTIANOVÁ, Antónia – VYSKOČILOVÁ, Jana. *Třetí vlna v kognitivně-behaviorální terapii*. Praha: Portál, 2018.

⁴⁷ KACZKURKIN, Antonia. Cognitive-behavioral therapy for anxiety disorders: an update on the empirical evidence, *Dialogues in Clinical Neuroscience*, 2015, s. 337-346.

asociace ve strukturách. Například tachykardie nevede k infarktu, přeplněné obchodní domy nevedou k násilnému útoku. Očekává se, že strach bude klesat konfrontací obávaného podnětu nebo reakcí a začleněním nápravných informací do paměti strachu. Expozice může mít několik forem, včetně imaginární in vitro, in vivo a interoceptivní. K imaginární expozici dochází, když si pacient živě představí obávanou situaci.

Další široce používanou metodou pro léčbu úzkostných poruch je kognitivní terapie. Ta je založena na Beckově třídílném modelu emocí, který uvádí, že myšlenky, pocity a chování spolu souvisí. V souladu s touto teorií se navrhuje změna maladaptivních myšlenek, aby se změnil maladaptivní afekty a chování pacienta. Kognitivní terapie se zaměřuje na zkreslené myšlenky pomocí řady technik, jako je identifikace nepřesného myšlení, zkoumání důkazů pro a proti automatickým myšlenkám, zpochybňování a změna maladaptivních myšlenek, změna problematického chování a vztahování se k jiným lidem adaptivnějším způsobem. Nedílnou součástí kognitivní terapie je psychoedukace o třídílném modelu emocí, různých formách zkresleného myšlení, např. myšlení typu všechno nebo nic, ukvapené závěry, diskvalifikace pozitivního a kognitivní restrukturalizace. Domácí úkoly jsou obvykle zadány tak, aby pacientům poskytly příležitost procvičit si tyto dovednosti v každodenním životě, což jim umožní osvojit si techniky, takže budou schopni aplikovat to, co se naučili, i po ukončení léčby. Při léčbě úzkostných poruch se kognitivní terapie nejčastěji používá ve spojení s behaviorálními technikami, které mohou zahrnovat expoziční cvičení. Kognitivní terapie je obvykle časově omezena na přibližně 20 sezení nebo méně a je zaměřena na problémy, které pacient identifikuje jako primární.

5. Etická reflexe specifických konverzačních nástrojů

5.1. Nástroje pro podporu péče o duševní zdraví

Konverzační nástroje umělé inteligence neboli Chatboty jsou softwarové programy, které jsou podle svých tvůrců navrženy tak, aby sloužily jako empatičtí a kompetentní partneři v konverzaci s uživatelem⁴⁸. Tyto programy využívají umělou inteligenci k tomu, aby byl uživatelům umožněn rozhovor podobný konverzaci se skutečnou osobou v reálném čase. Chatboty byly začleněny do některých aplikací pro podporu duševního zdraví, aby pomohly uživatelům pracovat s tím, jak se cítí, a to prostřednictvím otázek, hodnocení, návrhů aktivit a procvičování dovedností. Příklady aplikací řazené do této třídy chatbotů jsou Woebot, Wysa, Replika a Youper. Jejich detailnímu rozboru je věnována následující podkapitola. Některé chatboty pro podporu duševního zdraví jsou vytvořeny výzkumníky v oblasti psychologie, inženýry umělé inteligence a klinickými praktiky, jejichž odborné znalosti se spojily a tito specialisté vytvořili ve společném týmu příslušný produkt. Tyto chatboty působí na uživatele replikováním pomocných terapeutických technik kognitivně behaviorální terapie a terapie akceptace a závazku založených na klinických důkazech Evidence Based Therapy (EBT), jako jsou pozorné naslouchání, meditace, praktiky všímavosti a sledování nálady. Vědeckými metodami bylo prokázáno, že chatboty mohou být relativně účinným zdrojem pomoci při podpůrné léčbě pro zvládání některých symptomů úzkostných poruch. Specifická studie zkoumala, zda častější používání chatbota může vést k moderaci intenzity prožívání depresivních pocitů a nálad, a zjistila, že lidé užívající chatbota pro podporu duševní zdraví hlásili znatelný pokles intenzity depresivních pocitů ve srovnání s lidmi, kteří ho používali méně často⁴⁹. Další studie⁵⁰ porovnávala hodnoty úrovně symptomů deprese u lidí náhodně

⁴⁸ Vývojáři a provozovatelé těchto nástrojů často používají pro svoje produkty charakteristiky, které jsou z odborného psychologického hlediska mnohdy přinejmenším sporné. Tak je tomu například u pojmu empatie. Podrobnější rozbor takových případů bude uveden v podkapitole limitů umělé inteligence.

⁴⁹ INKSTER, Becky. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR mHealth and uHealth*, 2018, roč. 6, č. 11.

⁵⁰ FITZPATRICK, Kathleen. 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*, 2017 roč. 4, č. 2.

vybraných k používání chatbota Woebot na základě technik z kognitivně behaviorální terapie nebo stavu pouze na základě běžného poskytování popisných informací. Skupina, které byly poskytovány pouze popisné informace, měla za úkol navštívit informační platformu Národního institutu duševního zdraví, poskytující informace o metodách zvládnání úzkosti a deprese, a pročitat si příslušné informace. Na konci experimentu hlásila skupina vybraná pro užívání chatbotu významný pokles pocitů deprese a úzkosti, zatímco skupina vybavená pouhými informacemi žádnou změnu nezaregistrovala. Studie z recenzovaných zdrojů vesměs doporučují používání chatbotů z oblasti péče o duševní zdraví jako podpůrnou léčbu ke snížení pocitů deprese a úzkosti, avšak pod podmínkou, že se proces odehrává pod dohledem kliniků.

Aplikace pro podporu péče o duševní zdraví jsou stále rozšířenější a poměrně snadno dostupné pro každého, kdo má přístup k internetu a mobilnímu telefonu. Společnosti vyvíjející a provozující chatboty obvykle nezveřejňují počty uživatelů nicméně uvádějí, že například jeden z nejrozšířenějších chatbotů Woebot si každý týden vyměňuje s uživateli miliony zpráv⁵¹. Nedávné odhady naznačují, že je v současné době k dispozici mezi 10 000 až 15 000 veřejně dostupných aplikací, které je možné podle nabízené funkčnosti zařadit do skupiny podpůrných prostředků pro oblasti péče o duševní zdraví⁵². Rozvoj umělé inteligence v oblasti zdravotnictví byl urychlen zřetelným finančním zájmem investičních společností, co investují stovky milionů dolarů (USD) do vývoje aplikací v oblasti péče o duševní zdraví. Některé odhady uvádějí, že se finanční hodnota trhu v této oblasti v budoucích letech zmnohonásobí a dosáhne do roku 2028 120 miliard USD⁵³. Aby se mohl potenciální uživatel aplikací rozhodnout, zda je pro něj jejich používání tou správnou volbou, je důležité porozumět, jaké jsou přínosy a omezení aplikací pro duševní zdraví. Pro toto porozumění nabízí provozovatelé daných aplikací

⁵¹ Woebot CEO Michael Evers on AI in mental health, and how to get a chatbot to bond with a human. (30.7.2021) [2022-07-25]. < <https://vator.tv/news/2021-07-30-woebot-ceo-michael-evers-on-ai-in-mental-health-and-how-to-get-a-chatbot-to-bond-with-a-human>>.

⁵² NEARY, Martha. State of the Field of Mental Health Apps, *Cognitive and Behavioral Practice*, roč. 25, č. 4, 2018, s. 531-537.

⁵³ Woebot CEO Michael Evers on AI in mental health, and how to get a chatbot to bond with a human. (30.7.2021) [2022-07-25]. < <https://vator.tv/news/2021-07-30-woebot-ceo-michael-evers-on-ai-in-mental-health-and-how-to-get-a-chatbot-to-bond-with-a-human>>.

celou řadu vodítek, přičemž provozovatelé kladou velký důraz na fakt, že aplikace pro podporu péče o duševní zdraví neslouží ke stanovení diagnózy duševní poruchy nebo nemoci⁵⁴ a nenahrazují ani odbornou pomoc příslušného klinika. V prostředí Spojených států leží ovšem hlavní motiv těchto deklarací ve snaze být chráněn před případnými žalobami ze strany uživatelů, a to v případech, kdy dojde k ohrožení nebo zhoršení duševního zdraví uživatele aplikace. Pokud potřebuje jedinec léčebnou intervenci v oblasti duševního zdraví, měl by podle provozovatelů systémů umělé inteligence vždy navštívit zařízení poskytující příslušnou zdravotní péči. Některé aplikace nabízí i přímý kontakt na reálné terapeutky.

Jaké jsou nejčastěji uváděné potenciální přínosy aplikací užívajících umělou inteligenci? Především jsou aplikace snadno dostupné. Aplikace podpory péče o duševní zdraví jsou v průměru cenově dostupnější a jejich aktivace vyžaduje menší úsilí než tradiční terapie, které mohou požadovat zdravotní pojištění, osobní návštěvu a vyhledání terapeuta a komunikaci s ním. Aplikace jsou dostupné prostřednictvím chytrých telefonů a dalších mobilních zařízení, což znamená, že uživatelé mohou spouštět aplikace a poskytovat sami sobě péči podle své volby kdekoli a kdykoli. Kromě toho mnoho aplikací nabízí kreativní a zábavné způsoby, jak se věnovat péči o duševní zdraví. Aplikace mohou být použity maximálně diskrétně a anonymně. V tom může někdo spatřovat výhodu, protože jedním z důvodů, proč lidé nechtějí vyhledat pomoc pro léčbu svých psychických obtíží u psychoterapeuta nebo psychiatra, může být stigmatizace. Aplikace podpory péče o duševní zdraví se snaží zpřístupňovat používání léčby duševního zdraví prostřednictvím různých doplňkových vzdělávacích iniciativ. Pomocí aplikace je rovněž možné sledovat míru pohody v reálném čase a pomoci tak uživateli mapovat pocit pohody na reálné situace v jeho životě. Aplikace také usnadňují sledování dat a většinu času se uživatelé nemusí starat o to, aby si dělali poznámky nebo si pamatovali, jak se cítili před týdnem nebo měsícem nebo při nějaké specifické události. Mnoho aplikací sleduje i náladu a vytváří souhrny v aplikaci pokaždé, když dokončí uživatel relaci. Pro důvěryhodnost aplikací je podstatné, že existují důkazy, že jsou aplikace funkční a že výzkumné studie

⁵⁴ Tamtéž

publikované v impaktovaných zdrojích potvrzují funkčnost vybraných recenzovaných aplikací pro specifickou oblast péče o duševní zdraví⁵⁵.

S použitím aplikací jsou však spojeny i značné výzvy a určitá rizika. Ne každý má přístup k aplikacím, a i když se zdá, že přístup pomocí chytrého telefonu a internetu je stále běžnější, neplatí to pro každého potenciálního uživatele. Navíc většina populace ze specifických nejzranitelnějších skupin nemá většinou přístup žádný. Poskytování služeb podpory péče o duševního zdraví prostřednictvím chytrého telefonu by mohlo znamenat, že ti, kdo nemají přístup k technologii chytrých telefonů, nebudou mít ani přístup k těmto aplikacím. Kromě toho existuje jen málo aplikací, které nabízejí své služby v jiných jazycích než v angličtině. Uživatelé, kteří neovládají angličtinu, tak mohou mít při výběru aplikace pro podporu péče o duševní zdraví velmi omezené možnosti.

Aplikace pro podporu péče o duševní zdraví jsou nové formy péče ve srovnání s tradičními terapiemi a existuje mnoho vývojářů, kteří se snaží vytvářet nejrůznější produkty. Určitým paradoxem je například to, že léky musí v Česku schvalovat Státní ústav pro kontrolu léčiv dříve, než jsou dostupné v lékárnách pacientům. Stejně tak musí léčebné postupy používané ve zdravotnických zařízeních vyhovovat zásadě *Lege artis*, což není případ uvedených aplikací – stejná situace je ovšem i v ostatních zemích Evropské unie a ve Spojených státech. V důsledku toho není na vývojáře těchto aplikací kladen dostatečně účinný tlak, aby byly aplikace vytvořeny podle postupů založených na klinických důkazech. S tolika dostupnými aplikacemi na trhu má pravděpodobně jen malá část z nich použití založené na klinickém výzkumu a může proto být problém oddělit funkční aplikace od neúčinných nebo dokonce závadných. Existují i další limity mající silné etické konotace. Těm je věnována podkapitola etické limity nástrojů umělé inteligence pro péči o duševní zdraví.

⁵⁵ GRAHAM, Andrea. Coached mobile app platform for the treatment of depression and anxiety among primary care patients: a randomized clinical trial. *JAMA Psychiatry*, 2020, s. 906–914.

5.2. Popis vybraných chatbotů

Woebot je chatbot s umělou inteligencí využívající principy kognitivně-behaviorální terapie a podporující uživatele při zvládnání jejich úzkostných myšlenek a pocitů. Prostřednictvím instrukcí Woebot nejprve vyzve uživatele, aby specifikovali svou náladu a podrobnosti vysvětlující jejich náladu, a pak reaguje tím, že navrhuje nástroje, dovednosti a strategie, které uživateli pomohou náladu zvládat. Kromě zasílání zpráv si uživatelé mohou také prohlížet graf svých záznamů nálady v průběhu času a sledovat psychoedukační média. Uživatelé mají rovněž přístup k řízeným cvičením pro zvládnání negativních myšlenek o sobě a druhých metodou jejich zpochybnění nebo pro zvládnání škodlivého stresu. Na konci každé konverzace s Woebotem jsou uživatelé vyzváni, aby ohodnotili kvalitu konverzace. Je zároveň třeba zdůraznit, že Woebot není podle svých tvůrců určen k použití v případě nouze nebo ke zvládnání psychiatrických krizí.

Woebot, nabízený jako „*váš přátelský odborník na péči o sebe*“, je přístupný prostřednictvím mobilních zařízení. Aplikace obsahuje lekce, interaktivní cvičení a videa vycházející z vybraných prvků kognitivně behaviorální terapie. Interakce s Woebotem jsou navrženy tak, aby každá trvala zhruba 1-2 minuty. Doporučení pro použití Woebotu je zaměřeno na uživatele ve věku 13-19 let, kteří se potýkají se stresem a zdravotními potížemi, jako jsou problémy se spánkem, s úzkostí apod. Empirická podpora účinnosti pro Woebot vychází z randomizované kontrolované studie, jež ale zahrnovala pouze věkovou skupinu mladistvých⁵⁶. Aplikace se snaží zaujmout uživatele každý den pomocí pobídkových notifikací. Ti, kdo se podle tvůrců aplikace potýkají s mírným stresem nebo cítí mírnou úzkost, budou pravděpodobně mít z používání Woebota prospěch bez nutnosti intervence reálného psychoterapeuta. Naopak ti, kdo mají vážnější úroveň úzkostí, mohou mít z používání Woebota prospěch jen při předchozí návštěvě terapeuta. Ten indikuje vhodnost podpůrné terapie prostřednictvím chatbotu, s kterým procvičí potřebné dovednosti, a následně terapeuta instruuje, že se Woebot chová jako robot a nemůže vždy přesně reagovat na to, co mu uživatel říká. Obsah aplikace odráží některé základní

⁵⁶ FITZPATRICK, Kathleen. 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*, 2017 roč. 4, č. 2.

podpůrné procedury kognitivně-behaviorální terapie, jež má silnou důkazní základnu pro zvládnání úzkostných a depresivních stavů, a to i tehdy, kdy je poskytována v interakci mezi terapeutem a pacientem a za podpory vhodných technologických platforem⁵⁷. Aplikace také zahrnuje procvičování dovedností týkajících se vděčnosti a všímavosti, kdy se tvůrci aplikace inspirovali terapií přijetí a závazku⁵⁸. Aplikace byla hodnocena v randomizované kontrolované studii⁵⁹. Informace aplikace jsou prezentovány v různých formátech, včetně krátkého textu a videí, kdy je například k vidění minutové video „*Co je všímavost*“. Informace jsou poměrně stručné, zejména ve srovnání s tím, jak dlouho může trvat terapeutická vysvětlení konceptů kognitivně behaviorální terapie. Uživatel se rovněž může rozhodnout, zda chce další vysvětlení. Za zmínku také stojí, že většina interakcí s aplikací je navrstvena v ověřujících prohlášeních. Ta mají navozovat pocit empatie ze strany aplikace. Aplikace také často využívá jméno uživatele a tím její výroky působí osobněji. Opakovaně i vybízí uživatele, aby se po dokončení úkolů v aplikaci chválili. Woebot rovněž zahrnuje časté sledování nálady, například pomocí souboru otázek měřících příznaky deprese. V souladu s praxí kognitivně behaviorální terapie Woebot často vyhodnocuje změny nálady po aktivitách, jako je hluboké vědomé dýchání a relaxace. Informace poskytované uživatelům se zdají být jasné a vhodné pro populaci věkové skupiny mladistvých nezvládajících distres a majících obavu o duševní pohodu. Aplikace však jasně nespecifikuje, kdo by měl aplikaci používat a jaké nástroje nebo typy interakcí s Woebotem by mohly být nejlepší na základě toho, co se děje v životě uživatele. Někteří uživatelé proto nemusejí mít zcela jasno v tom, zda aplikaci používají takovým způsobem, které by byl pro ně nejužitečnější, nebo zda potřebují vyšší úroveň péče poskytovanou psychoterapeutem nebo psychiatrem. Dále, jak aplikace informuje uživatele, když ji poprvé otevřou: žádný člověk nesleduje průběh živé konverzace mezi uživatelem a chatbotem. To znamená, že když Woebot umožňuje uživatelům zadat volný text, tj. napsat svou vlastní odpověď, tak poskytnutá zpětná vazba může být nesouhlasná

⁵⁷ HOFMANN, Stefan. The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research*, 2012, s. 427-440.

⁵⁸ ŠLEPECKÝ, Miloš - PRAŠKO, Ján – KOTIANOVÁ, Antónia – VYSKOČILOVÁ, Jana. *Třetí vlna v kognitivně-behaviorální terapii*. Praha: Portál, 2018.

⁵⁹ National Institute of Mental Health. Depression in College Students. (2017) [2022-5-31]. <<https://www.nimh.nih.gov/health/publications/depression-and-college-students/index.shtml> > .

nebo neúplná. Například při učení a procvičování dovednosti náročnějšího myšlení založeného na kognitivně behaviorální terapii nemusí vždy Woebot reagovat na tuto myšlenku, a zpětná vazba, kterou poskytne, nemusí být příliš vhodná pro typ výzvy, již by myšlenka potřebovala z perspektivy kognitivně behaviorální terapie poskytované psychoterapeutem nebo psychiatrem. Při osobním testování aplikace jsem dospěl k jednoznačnému závěru, že je Woebot koncipován spíše pro efektivní zvládnání kognitivně jednodušších dialogů.

Youper je konverzační robot užívající umělou inteligenci navržený tak, aby uživatelům pomáhal identifikovat, sledovat a zpracovávat jejich myšlenky a pocity. Využitím kognitivně behaviorální terapie a její třetí generace terapie akceptace a závazku Acceptance Commitment Therapy (ACT)⁶⁰ a všímavosti může uživatelům pomoci zlepšit náladu a spánkové vzorce a snížit příznaky úzkosti, stresu a deprese. Aplikace obsahuje čtyři hlavní funkce: chat, deníkové záznamy, záznamy nálady a hodnocení emočního zdraví. Aplikace vybízí uživatele, aby popsali, jak se aktuálně cítí, s možností požádat aplikaci o podporu při vyplnění údajů, a také žádá, aby uživatelé identifikovali faktory nebo události, které přispěly k tomu, jak se aktuálně cítí. Nakonec aplikace navrhuje několik aktivit zlepšujících prožití zbytku dne. Protokol deníku zobrazuje všechny reakce uživatele na otázku *Jak se cítíš?* Funkce Mood Log shromažďuje odpovědi uživatelů na otázku *Proč jste se tak cítil?* Doporučený věk uživatele je 12+. Aplikace je zajímavá tím, že je založena na sledování příznaků. Sledovače příznaků jsou navrženy tak, aby sledovaly symptomy i další životní události, které mohou ovlivnit to, jak se uživatel cítí. Sledovače symptomů lze použít ke sledování závažnosti symptomů, emocionálních nebo environmentálních spouštěčů, časných známek relapsu. Sledování příznaků lze použít k poskytování informací klinikovi, ke snížení rizika zmeškaných diagnóz a zajištění správné léčby. Pokud jsou sledovány příznaky, může mít pacient přesný záznam o tom, jak se jeho psychický stav vyvíjel, a může jej sdílet se svým lékařem a poskytnout mu důležité informace. Sledování příznaků lze využít i v rovině osobního použití, tj. u těch,

⁶⁰ ŠLEPECKÝ, Miloš - PRAŠKO, Ján – KOTIANOVÁ, Antónia – VYSKOČILOVÁ, Jana. *Třetí vlna v kognitivně-behaviorální terapii*. Praha: Portál, 2018.

kdo chtějí pečlivě sledovat své vlastní příznaky a identifikovat, co může pomoci nebo i spouštět špatnou náladu, úzkost atd. Lidé tradičně sledovali své příznaky metodou tužky a papíru. Použití aplikací nebo online programů ke sledování je spolehlivější a pohodlnější řešení. Uživatelé sledující svůj duševní stav si mohou vybrat ze seznamu příznaků. Mnoho sledovacích aplikací má výhodu také v tom, že uživatelům umožňuje nastavit si připomenutí, po němž se uživatelům každý den zobrazí výzva, že mají zaznamenat příznaky, což je úkol, který lze jinak snadno zapomenout nebo přehlédnout. Jedna studie zkoumala výhody monitorování nálady, stresu a zvládání pomocí aplikace pro chytré telefony⁶¹, s jejíž pomocí se mladí lidé ve věku 14–24 let – v obavách o své duševní zdraví - sami monitorovali po dobu 2–4 týdnů. Studie zjistila, že fakt samotného sebezpozorování pomohl zvýšit emocionální sebeuvědomění vedoucí ke snížení příznaků deprese. Je třeba brát v úvahu i to, že neexistuje žádný odborný článek publikovaný v impaktovaném periodiku, který by popisoval výzkum účinnosti Youper AI za použití vědeckých metod.

Replika AI je konverzační robot využívající umělou inteligenci, jež svým uživatelům umožňuje vést si digitální deník prostřednictvím konverzace s aplikací. Umělá inteligence se učí a rozvíjí svoje algoritmy na základě nových dat zadaných uživatelem. Uživatelé aktualizují znalostní databázi umělé inteligence přidáním fotografie a nahráním různých dalších datových objektů, jako je digitální záznam hlasu. V systému je nabízen program cesty, kde se uživatel poutavým způsobem účastní několika misí zahrnujících aspekty sledování nálady, sebezpozorování, psychoedukaci o náladě a dovednostech zvládání a stanovování cílů. Plněním misí se získávají specifické dovednosti, které pak může uživatel použít v reálných situacích. Aplikace také umožňuje uživatelům sledovat jejich pokrok v dosahování určitých osobních cílů, jako je snížení stresu a úzkosti a prožívání pocitu euforie. Pokrok je indikován zobrazením statistik, zobrazením vývoje znalostní databáze uživatele a sledováním různých akcí nebo chování uživatele. Výzkum této aplikace vycházel z předchozích analýz, které naznačovaly, že podobné typy aplikace

⁶¹ KAUER, Silvia. Self-monitoring Using Mobile Phones in the Early Stages of Adolescent Depression: Randomized Controlled Trial. *J Med Internet Res* 2012.

mohou být slibným zdrojem sociální podpory pro lidi s úzkostnými stavy spouštěnými ve specifických sociálních situacích. Většina tohoto výzkumu však byla provedena v kontextu intervencí sociální podpory, které se zabývají stresovými situacemi nebo problematikou zlepšování zdraví. Rozsahem malý výzkum zkoumal sociální podporu získanou od aplikací v každodenních kontextech. Vzhledem k tomu, že sociální opora se projevuje nejen v krizích, ale i v každodenních situacích, a že každodenní sociální opora tvoří základ podpory při stresujících událostech, zaměřil se výzkum na zkoumání typů každodenní sociální opory, které lze získat od podobných aplikací. Výsledek ukázal, že Replika AI poskytuje uživateli, se kterým komunikuje, určitou úroveň sociální aktivity, jež může pomoci omezit osamělost, poskytuje prostor, kde se může uživatel cítit bezpečně a kde mohou uživatelé diskutovat o jakémkoli tématu beze strachu z odsudku nebo zranění. Zvyšuje také pozitivní vliv prostřednictvím povzbuzujících a chválících zpráv a poskytuje uživateli užitečné informace⁶².

5.3. Etické problémy a výzvy při vývoji a užití chatbotů

Použití konverzačních aplikací – chatbotů představuje pro společnost příležitost poskytovat podporu při řešení psychických problémů i lidem, kteří by se k ní jinak nedostali. Navzdory možným výhodám konverzačních nástrojů umělé inteligence však existují také významné etické výzvy spojené s jejich použitím. Pro etické použití těchto prostředků umělé inteligence je proto nutné identifikovat rizika s tím spojená a zajistit zavedení takových opatření, která tato rizika budou buď eliminovat anebo přinejmenším minimalizovat.

Jedním z etických problémů, který je přímo spojen s designem chatbotu je problém možné předsudečnosti⁶³. Chování chatbotů může být náchylné k předsudkům již v designu, například nevědomým upřednostňováním určitého rasového nebo etnického

⁶² GRIFFITH, Caroline. A User Experiences of Social Support From Companion Chatbots in Everyday Contexts: Thematic Analysis *J Med Internet Res* 2020.

⁶³ FISKE, Amelie. A Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy, *J Med Internet Res* 2019.

původu⁶⁴. K předsudečnosti je náchylná i znalostní databáze používaná k trénování algoritmů strojového učení a způsobů, jakými konverzační agent generuje nové znalosti, přičemž tento typ zkreslení není tak zjevný jako fyzický vzhled nebo jazykové zkreslení. Algoritmické zkreslení nastává, když softwarový program dělá systematické chyby, které vytvářejí nespravedlivé výsledky, jako je upřednostňování jedné skupiny uživatelů před ostatními. Zdroje zkreslení v souborech dat používaných k vývoji konverzačních agentů mohou zahrnovat problémy s chybějícími údaji, chybnou klasifikací a chybou měření a malou velikostí vzorků, což vede k podcenění a nepřesným předpovědím pro specifické podskupiny uživatelů. Kromě toho mohou implicitní charakteristiky programátorů a organizací, které sbírají, vybírají nebo používají data k trénování algoritmů, také představovat předsudečná zkreslení. Aplikace konverzačních agentů v různých zemích a různých kulturách je zvláště náchylná k těmto předsudkům v designu, pokud organizace nasazující technologii adekvátně nezváží demografické charakteristiky a specifické potřeby cílových koncových uživatelů. Vývojáři konverzačních agentů by měli vzít v úvahu potenciální zkreslení při navrhování a testování konverzačních agentů. Zásadní je proto zahrnutí dat od cílové populace a jejích různých komunit, které by mohlo být ovlivněno rozdíly v návrhu, testování a implementaci těchto technologií. Neustálý výzkum a vyhodnocování konverzačních činitelů jsou nezbytné k tomu, aby se zabránilo zkreslení a nedostatkům v datech používaných algoritmy strojového učení, které by mohly přispět k socioekonomickým rozdílům v poskytování služeb.

Dalším etickým problémem chatbotu je riziko možného poškození jeho uživatele. Protože konverzační agenti fungují s určitou úrovní autonomie, existuje potenciální riziko poškození uživatelů, pokud technologie adekvátně neřeší scénáře, kdy systém detekuje potenciální nebo bezprostřední bezpečnostní rizika. Například v případě osoby, která konverzuje s chatbotem, by mohla naznačit, že má sebevražedné myšlenky a má plány okamžitě zemřít, chatbot by nemusel v takovém případě vhodně a včas zareagovat. Někteří pacienti, například pacienti s významnými psychotickými symptomy nebo kognitivními deficity, také nemusí být vhodnými kandidáty pro použití konverzačních

⁶⁴ LUXTON, David D. Ethical implications of conversational agents in global public health. *Bulletin of the World Health Organization* 2020, s. 285 - 287.

prostředků, protože většina populací, které se používají pro strojové učení chatbotů, nemá tyto specifické symptomy. Není proto jednoznačné, jak by aplikace reagovala v případě, kdy v důsledku psychotické ataky uživatele hrozí riziko jeho sebepoškození nebo násilí vůči jeho okolí. Pro vyřešení těchto problémů by správci technologie měli mít spolehlivý postup pro zjištění zamýšleného použití aplikace a pro prověření vhodnosti uživatelů, než jim bude udělen přístup ke konverzačnímu agentovi. Technologie by také měla být navržena tak, aby byla schopna automaticky monitorovat rizika a poté přijmout vhodná opatření. V některých scénářích může systém konverzačních agentů poskytnout okamžité zobrazení zdrojů pomoci, jako je linka krizové intervence nebo jiné užitečné zdroje. V jiných situacích jsou potřebné postupy, například aby ošetřující lékař mohl přezkoumat informace z komunikace mezi pacientem a aplikací, a pokud by to bylo vhodné, kontaktoval uživatele, aby mu poskytl vhodnou pomoc. V současné době většina společností, vyrábějících konverzační prostředky pro použití souvisejících s péčí o psychické zdraví, explicitně deklaruje, že je prodává veřejnosti za účelem poskytování informací nebo cvičení dovedností, a nikoli jako náhradu za léčbu klinikem. Proto mnoho etických a právních požadavků, které by přitom měli zdravotničtí pracovníci dodržovat, například povinnost varovat v případě podezření, že se stal kriminální delikt, není bráno tvůrci a provozovateli systémů umělé inteligence na zřetel. Je proto nezbytné, aby správci těchto systémů poskytovali uživatelům dostatečné informace o rozsahu použití, rizicích, limitech a očekáváních ohledně používání konverzačních agentů. Kromě toho musí správci aplikací identifikovat a posoudit přiměřenost zdrojů, jako je krizová podpora nebo všeobecné zdravotnické služby, dostupné v zemích a regionech, kde bude technologie používána.

Soukromí uživatelů je další etický problém, který – pokud by nebyl náležitě řešen – by potenciálně mohl poškodit uživatele konverzačních agentů. V průběhu konverzace s lidmi mohou konverzační agenti shromažďovat velké množství soukromých a citlivých informací. Podobně jako při používání online zdravotních služeb se mohou lišit zákony a předpisy týkající se zajištění soukromí uživatelů napříč mezinárodními hranicemi. Správci konverzačních agentů by si měli být vědomi toho, jaké jsou místní požadavky, pokud existují, a informovat koncové uživatele o omezeních soukromí a ochranných opatřeních zavedených k zabezpečení dat uživatelů.

Mezi další etické problémy chatbotů patří riziko porušení principu spravedlnosti v přístupu k technologii. Uvědomění si potenciálu konverzačních agentů ve

zdravotnických službách závisí na tom, zda jsou tyto technologie dostupné populacím, které by z nich mohly mít prospěch. Nedostatek investic do technologické infrastruktury v nedostatečně obsluhovaných komunitách a zemích, stejně jako náklady na ni, jsou hlavním omezením pro přijetí jakékoli pokročilé technologie. Nedostatek vzdělání a nízká technologická gramotnost jsou další problémy, které by mohly omezit přístup lidí ke konverzačním prostředkům a jejich používání. To má za následek větší rozdíly ve zdravotní péči mezi populacemi. Aby byli konverzační agenti životaschopní, musí správci technologie zvážit přístup uživatelů k technologii a obeznámenost s jejím používáním. Stejně jako u tradiční telemedicíny používající videokonference, telefonické konzultace a vzdálené monitorování mohou omezení technologie nebo preference koncových uživatelů způsobit, že použití konverzačních agentů bude nevhodné. V tomto případě by správci měli zvážit alternativy, jako jsou osobní služby.

Dalším problémem spojeným s provozováním chatbotů jsou nedostatečně systematicky řešené etické výzvy. Zatímco konverzační agenti mají potenciál pomoci řešit některé potřeby zdravotních služeb po celém světě, musí vývojáři a poskytovatelé služeb konverzačních agentů zvážit bezpečnost, důstojnost a respekt uživatelů, aby zajistili etické použití a aplikaci technologie. Jeden ze způsobů, jak začít řešit současné nově vznikající etické výzvy, spojené s používáním konverzačních agentů, je revidovat nebo vyvinout nové profesionální etické kodexy a praktické pokyny. Jak bylo uvedeno ve třetí kapitole této práce, nezabývá se většina současných etických kodexů a praktických pokynů ve zdravotnictví⁶⁵ používáním technologií simulujících a nahrazujících lidské profesionály. Světová zdravotnická organizace by mohla vytvořit kooperativní mezinárodní pracovní skupinu, která by přezkoumala stávající etické zásady a pokyny a vydala celosvětové doporučení k zajištění etického používání nástrojů založených na umělé inteligenci, včetně konverzačních agentů a souvisejících technologií. Žádoucího výsledku by se nejlépe dosáhlo prostřednictvím spolupráce mezi všemi zúčastněnými stranami, včetně zástupců osob, které jsou nedostatečně obsluhovány a nejvíce postiženy rozdíly ve zdravotní péči. Společná obhajoba vhodného

⁶⁵ Výjimku tvoří iniciativa Evropské unie, jak bylo popsáno ve třetí kapitole.

použití a výhod konverzačních agentů by také pomohla vyřešit obavy z technologií a potenciálně urychlit její přijetí ve společnosti.

5.4. Limity prostředků péče o duševní zdraví mající etický přesah

Jedním z hlavních limitů prostředků péče o duševní zdraví využívajících umělou inteligenci je absence reálného vztahu mezi pacientem a terapeutem, jenž aplikace pouze simuluje. Vztah mezi pacientem a terapeutem je přitom jeden z hlavních léčebných faktorů psychoterapie⁶⁶, nazývaný také terapeutická aliance. Termín je používán k popisu profesionálního vztahu mezi poskytovatelem péče, tj. psychoterapeutem, a pacientem, kdy psychoterapeut doufá, že se podaří tuto alianci vytvořit a že společně s pacientem dosáhnou terapeutického cíle. Terapeutický vztah je klíčovým společným faktorem spojeným s žádoucími léčebnými výsledky nezávislými na konkrétním typu léčby. Jinými slovy, kvalita terapeutického vztahu je klíčovým determinantem toho, že léčba psychických problémů pacienta postupuje přinejmenším tak, jak to předpokládá indikovaná léčebná metoda. Zkoumáním psychoterapeutického vztahu z perspektivy prostředků umělé inteligence se věnuje řada studií⁶⁷, kladoucích celou řadu závažných otázek a doporučení, jak příslušná rizika minimalizovat prostřednictvím činností zdravotních regulátorů a etických kodexů. Vztah mezi poskytovatelem profesionální péče a pacienty není běžným společenským vztahem. V této oblasti existují důležité právní a etické povinnosti, neboť poskytovatelé péče mají ve vztahu k pacientům silnou pozici a existuje proto určité riziko poškození a zneužívání tohoto postavení pacientů poskytovatelem péče. Etické závazky mezi poskytovatelem péče a pacienty jsou ústředním bodem etických kodexů profesí v oblasti péče o duševní zdraví. Etický metakód Evropské federace psychologických asociací EFPA například stanovuje: „Profesionální chování psychologa musí být posuzováno v rámci jeho odborné role, charakterizované

⁶⁶ PROCHASKA, James – NORCROSS, John. *Psychoterapeutické systémy*. Praha: Grada 1996. s. 19.

⁶⁷ LUXTON, David. Recommendations for the ethical use and design of artificial intelligent care providers, *Artificial Intelligence in Medicine* 2014, s. 1-10,

profesionálním vztahem“⁶⁸. Tyto požadavky se stávají obzvláště důležité kvůli soukromé, vysoce osobní a někdy silně emocionální povaze vztahu navázaného s psychoterapeutem. *Ústředním bodem terapeutického vztahu je důvěra, protože ta tvoří základ pro vztah mezi poskytovatelem péče a pacientem. Aby bylo možné stanovit přesnou diagnózu a poskytnout optimální léčbu, musí být pacient schopen sdělit všechny relevantní klinické informace poskytovateli péče. Stejně jako mezi lidmi, poskytovateli terapeutické péče a jejich pacienty, bude důvěra také kritickou proměnnou pro pacienty, kteří interagují s prostředky umělé inteligence. Problém důvěry vyvolává důležitou otázku, zda je či není etické simulovat chování člověka prostředkem umělé inteligence natolik, že lidé věří, že komunikují s člověkem. Domnívám se, že to etické není, protože je takto získaná důvěra založená na klamu, a navíc z povahy věci plyne, že aby byla důvěra vědomá a autentická, musí být reciproční, což u dané aplikace nelze dosáhnout, protože tuto charakteristiku může naplňovat pouze člověk.* Jiný názor zastává německý profesor informatiky Weizenbaum, jenž roku 1966 prezentoval poměrně jednoduchý program ELIZA, zpracovávající vstup od uživatele psaný přirozeným jazykem. Přestože program pouze vybíral z uživatelských odpovědí klíčová slova a z nich tvořil další otázky, podleho mnoho uživatelů dojmu, že program přemýšlí, nebo že jim odpovídá člověk⁶⁹. Autor programu poznamenává, že ELIZA lidi skutečně klamala, ale argumentuje, že: *“ELIZA vytvořila v myslích mnoha lidí, kteří s ní mluvili, tu nejpozoruhodnější iluzi porozumění. Lidé, kteří velmi dobře věděli, že konverzují se strojem, na to brzy zapomněli. Tato iluze byla obzvláště silná a nejhouževnatěji se držela mezi lidmi, co o počítačích věděli jen málo nebo vůbec nic. Často požadovali, aby jim bylo dovoleno konverzovat se systémem dál i v soukromí a po nějakém rozhovoru s ním i přes moje vysvětlení trvali na tom, že jim stroj skutečně rozuměl“*⁷⁰. Weizenbaum poznamenal, že lidé, co mluvili s ELIZOU, při tom neustupovali od nároků na skutečnost, přičemž uvedl, že i když tito lidé vědí, že

⁶⁸ LINNDSAY *Etika pro evropské psychology*. Praha: Triton 2010. s. 243.

⁶⁹ ELIZA - A Computer Program for the Study of Natural Language Communication between Man and Machine, *Communications of the Association for Computing Machinery* 1966. s. 36-45.

⁷⁰ WEIZENBAUM, John. *Computer power and human reason: from judgment to calculation*. San Francisco: W.H. Freeman. (1976) [2022-05-31]. <https://www.researchgate.net/publication/286058724_Computer_Power_and_Human_Reason_From_Judgment_to_Calculation>.

konverzují se softwarovým programem, tak stále věří, že simulovaný psychoterapeut je skutečný, protože jeho osoba přiřazuje význam a interpretaci tomu, co říká ELIZA, což potvrzuje hypotézu dané osoby, že jim systém rozumí.

Klíčem k vytvoření a udržení terapeutického vztahu je empatické porozumění⁷¹. To pacientovi naznačuje a potvrzuje, že poskytovatel péče rozumí tomu, co pacient cítí a prožívá, a zajímá ho. Reflexe popisuje proces, jímž poskytovatel péče sděluje, že je pacient vnímán a přijímán a že zároveň reaguje na pocity pacienta. ELIZA svým základním textovým rozhraním a několika stovkami řádků počítačového kódu prokázala, že reflektovat pocity člověka, projevovat empatii a vtáhnout člověka do interakce není verbálně příliš komplikované. Současné systémy, jako jsou například chatboty, o nichž bylo pojednáno v předchozích kapitolách, jsou v tomto ohledu mnohem pokročilejší. Poskytovatelé simulované péče mohou verbálně reflektovat důležitá prohlášení vyjádřená pacientem, říkat všeobecně známé psychoterapeutické „hmmm“ a provádět pohyby modelem lidského obličeje nebo těla, reflektující pozornost a porozumění tomu, co pacient říká nebo vyjadřuje pohybem těla. Hlavní motivací těchto verbálních a neverbálních simulací je, aby chatbot působil na pacienta empatickým dojmem, jenž je důležitým faktorem pro vytváření a prohlubování vztahu. Řada výzkumů prokázala, jak významně je vjem empaticnosti konverzujícího chatbota závislý na kombinaci užitých slov, resp. vzorců vět, jimiž chatbot reaguje na uživatele⁷².

Z výzkumu je také známo, že si lidé mohou vytvořit silné emocionální vazby k technologickým zařízením a simulacím, jako jsou virtuální avataři⁷³ nebo sociální roboti⁷⁴. Můžeme tak očekávat, že pokud se lidé svěří prostředkům umělé inteligence, mohou je pak někteří prožívat jako skutečné osoby a vytvoří si k nim proto silné intimní

⁷¹ ROGERS Carl. *Teorie terapie a osobnosti*. Praha: Portál. 2020. s.185.

⁷² RASKIN, Hannah. Towards Empathetic Open-domain Conversation Models: A New Benchmark and Dataset. *Association for Computational Linguistics*. 2019. s. 5370–5381.

⁷³ COLLINS, Emily. Drawing parallels in human-other interactions: a trans-disciplinary approach to developing human-robot interaction methodologies. *Philos Trans R Soc Lond B Biol Sci*. 2019.

⁷⁴ FRONEMANN, Nora. Should my robot know what's best for me? Human–robot interaction between user experience and ethical design. *AI & SOCIETY*. 2022. s. 517-533.

vazby. I když si klient nebo pacient uvědomuje, že simulace poskytovatele léčebné péče je vytvářena strojem, lze očekávat, že během interakce zažije některý z nich intenzivní emoce. V takovém případě je na klinikoví, aby posoudil, zda je užívání prostředku umělé inteligence indikováno pro danou situaci nebo je naopak nežádoucí. Může se například jednat o přenos, kterým je proces, při němž pacienti nevědomě přesměrovávají pocity směrem k psychoterapeutovi symbolizujícímu významnou osobu v pacientově minulém nebo současném životě. I když přenos může být žádoucí a může mít v některých kontextech terapeutické důvody, je důležité také zvážit, zda může být i nežádoucí.

Potenciál iatrogenních účinků prostředků umělé inteligence lze přirovnat ke stejným typům etických problémů a rizik identifikovaných při používání klinické virtuální reality, včetně exacerbace nebo indukce symptomů duševního zdraví, které dříve nebyly u pacienta přítomny. V tomto případě však zůstává nedořešen etický problém terapeutického vztahu, který je dán tím, že je pacient uváděn v omyl. Z hlediska psychologického může být postup účinný, ale z hlediska etického je nesprávný – zde je možné uplatnit základní poučku bioetiky, že co je možné není automaticky eticky ospravedlnitelné. Tato problematika představuje závažný etický problém, zejména při práci se zranitelnými osobami. Bioetika a etika psychoterapie je zaměřena na podporu rozhodování, zda je specifický přístup v oblasti léčby s ohledem na základní etické principy správný či nikoliv. Z tohoto pohledu jednoznačně plyne, že pokud by prostředek umělé inteligence nahrazoval psychologa a pokud by klamně vytvářel podmínky pro vznik a rozvoj psychoterapeutického vztahu, byl by porušen základní etický problém nonmaleficence a takový postup by byl nesprávný. Příčiny poškození pacienta mohou být v tomto kontextu různorodé. Může to být chybná diagnóza, kterou v sobě prostředek implicitně obsahuje, tvrzení poskytovatelů, že prostředek diagnózu neprovádí⁷⁵, nevhodně zvolená terapie či správně zvolený, leč nesprávně provedený terapeutický zákrok.

⁷⁵ Woebot CEO Michael Evers on AI in mental health, and how to get a chatbot to bond with a human. (30.7.2021) [2022-07-25]. < <https://vator.tv/news/2021-07-30-woebot-ceo-michael-evers-on-ai-in-mental-health-and-how-to-get-a-chatbot-to-bond-with-a-human>>.

Dalším iatrogenním poškozením může být depersonalizace pacienta nebo klienta, kdy dochází k degradaci pacienta nebo klienta na objekt, se kterým se vyměňují data. Pacient se může nalézat v jedinečné životní situaci, do níž byl duševní poruchou nebo zatěžujícími symptomy uvržen. *Porozumět jí nemusí být triviální, neboť klade na psychoterapeuta nárok na individualizující vnímavost a porozumění. Takových nároků nemůže prostředek využívající umělé inteligence z principu dostát, protože postupy těchto prostředků vycházejí z nomotetického pojetí poznání osobnosti pacienta a individualizující přístup vyžaduje naopak idiografický přístup*⁷⁶.

Další riziko představuje nesprávná práce s emocemi. K naplnění etického principu nonmaleficence je povinností nejen vytvořit bezpečné terapeutické prostředí pro pacienty a jejich autentické prožívání emocí, ale i pozorovat emoční reakce pacientů a zmírňovat potenciálně zraňující emoční interakce klinicky vhodným způsobem. Nekontrolované systémy umělé inteligence, které nejsou schopny detekovat zhoršení symptomů v oblasti prožívání a projevoování emocí, představují významné iatrogenním riziko. Systémy umělé inteligence využívané pro péči o duševní zdraví musí být navrženy tak, aby ve vhodném okamžiku a ve spolupráci s pacientem byla dokončena interakce způsobem, který nezpůsobuje úzkost.

Specifickou oblast tvoří naplňování etického principu beneficence. Zvažování a uplatňování tohoto etického principu je podobně jako v medicíně složitější než aplikace jiného základního principu maleficence, protože určit, co je dobré pro pacienta a nalézt zároveň konsensus se samotným pacientem, nemusí být triviální úkol. Podobně jako v medicíně, kde je při společném hledání dobra pro pacienta nutné hledat optimum ve čtyřech oblastech bio-psycho-socio-spirituální⁷⁷, je vhodné i pro psychoterapii postupovat obdobně. Základním problémem oproti medicíně je fakt, že je v psychoterapii

⁷⁶ Problematice vztahu idiografického a nomotetického se v psychologické diagnostice věnoval americký psycholog Theodore Millon, který doporučoval, aby se při zjišťování a diagnostice psychických poruch začalo s nomotetickou perspektivou a po úvodní identifikaci poruchy se pokračovalo v diagnostice s idiografickou perspektivou.

⁷⁷ PŘÍHODA, Petr. Etika : Čtvero základních principů lékařské etiky [online]. [cit. 2011-12-23].

<<http://www.lf2.cuni.cz/ustav-lekarske-etiky-a-humanitnich-zakladu-mediciny-2-lf-uk/etika>>.

velké množství legitimních směrů, které se v pojetí toho, co je pro pacienta optimální, mohou značně lišit⁷⁸.

Například z kognitivně behaviorálního pohledu je primární charakteristikou člověka jeho chování, které je ovlivňováno tím, jak prostředí člověk vnímá a jak ho interpretuje. Zdrojem patologie jsou naučené chybné kognitivní konstrukce zkreslující vnímanou skutečnost a nevhodné kognitivní strategie pro zvládnání životních situací. V terapii se klient učí rozpoznávat škodlivost negativních myšlenek a experimentovat s chováním tak, aby došlo ke korektivní změně myšlení a chování.

5.5.Limity chatbotů z perspektivy etiky dialogického personalismu

V oblasti psychoterapie je velké množství směrů, které mají nejenom zcela odlišná východiska v pojetí péče o terapii přijímajícího jedince, ale mají i různorodá etická východiska. Zcela odlišný od kognitivně behaviorálního je ku příkladu přístup daseinsanalytické psychoterapie. Psychoterapeuti vycházející z daseinsanalytického směru se dívají na pojetí osobnosti jako množiny neměnných rysů kriticky. Pro ně je člověk obrazem jeho bytí ve světě. Svět lidí je osobní konstrukcí závislou na míře konvenčnosti jedince. Dasein terapie není zaměřena na universální pochopení světa, nýbrž na pochopení světa jednotlivce tak, jak ho on sám konstruuje a prožívá. Když jsme s druhými, tak víme, že jsou to vědomé bytosti, které nás budou hodnotit a soudit. To může vyvolat strach a pak se chováme tak, abychom byli viděni v lepším světle – bytí je redukováno na bytí pro druhé. Základem patologie je lhaní a sebeklamy člověka v různých úrovních života. Cílem terapie je za pomoci terapeutického dialogu dosáhnout autentičnosti a z perspektivy principů bioetiky obsáhnout možnou autonomii. Zvyšování vědomí se tak stává nejdůležitějším procesem, pomocí kterých si lidé uvědomí aspekty světa, které jim byly díky sebeklamu zakryty. K Dasein terapii je komplementární terapie aplikující dialogický přístup⁷⁹ indikovaný mimo jiné i pro úzkostné poruchy. Tento

⁷⁸ PROCHASKA, James – NORCROSS, John. *Psychoterapeutické systémy*. Praha: Grada 1996.

⁷⁹ SEIKKULA, Jaakko. Healing elements of therapeutic conversation: Dialogue as an embodiment of love. *Family Process*, 2005, roč. 44, č.4, s 461–475.

terapeutický koncept byl rozvíjen jako přístup, který u příjemce terapie klade na první místo jejich vlastní sílu k jednání a přivádí do léčby pacienta další relevantní osoby, aby se podíleli na potřebné pomoci. V centru pozornosti terapie jsou etická témata spojená s principem autonomie týkající se rozhodování. Jedna ze základních otázek mající přesah do etiky je, jaké volby se snažíme dělat, abychom u druhých lidí respektovali jejich vlastní sílu k jednání. Teoretická východiska tohoto psychoterapeutického směru vycházejí z dialogického personalismu rozlišujícího tři roviny dialogu - dialog jako techné, jako epistémé a jako mysterion⁸⁰. Na pokrok v terapii tohoto typu má zásadní vliv míra dialogičnosti, která je určena vnitřní spoluúčasti osob v dialogu. Takový dialog je vztahovým děním, které není redukováno tak, jak je tomu u rozhovoru s chatbotem, na přenos informací, ale týká se lidí, jako osob, tedy nezpředmětnitelných a nezastupitelných jedinců se svojí osobitou odpovědností a svobodou. Cílem takto pojímaného dialogu je porozumění a konsensuální jednání ve vzájemném spolubytí. Z hlediska dialogického personalismu nemůže být, v důsledku povahy chatbota, nastolen dialog mezi příjemcem psychoterapeutické péče a chatbotem – je to prostá výměna informací. Rozhovor s chatbotem není dialog jako techné, protože vlastnost chatbotu – bezvztahovost, je jeho základní překážkou. Z perspektivy dialogu jako techné nesplňuje chatbot základní podmínku odpovědnosti z důvodu nevztahovosti a ani dialogickou toleranci založenou na schopnosti sebezpochybnění aktérů dialogu. Rozhovor mezi příjemcem terapie a chatbotem není ani dialogem jako epistémé, ani jako mysterion. Takový rozhovor nevede k dialogickému poznávání, pro které je typické, že se neomezuje na vědění, které potvrzuje samo sebe, ale jde mu o vědění potvrzující samu skutečnost. Z perspektivy dialogu jako mysterion, ve které je klíčový pojem osoby, nenaplnuje takový rozhovor požadavek na nepoužitelnost osob, která je v současném dialogickém personalismu pojímána prostřednictvím vztahových charakteristik⁸¹. V takovém dialogu se osoby recipročně respektují a podporují ne jako prostředek, ale jako účel. *Je zjevné, že dialogický vztah je obsáhlejší než sociální vztah, na základě, kterého jsou chatboti koncipováni a proto z této etické perspektivy nemůže být dialog s chatbotem terapeutický.*

⁸⁰ POLÁKOVÁ, Jolana. *Smysl dialogu*. Praha: Vyšehrad 2008. s. 11-67.

⁸¹ VRÁNA, Karel. *Dialogický personalismus*. Praha: Zvon 1996. s. 22.

6. Prevence etických problémů umělé inteligence

6.1. Syntéza etických principů

Cílem této kapitoly je hodnotit etické a sociální důsledky převádění aplikací umělé inteligence do péče o duševní zdraví napříč obory psychiatrie, psychologie a psychoterapie. Výzkum umělé inteligence má rostoucí klinický význam pro terapeutické aplikace zaměřené na péči o duševní zdraví. S inovacemi od virtuálních psychoterapeutů přes sociální roboty v péči o demenci a autismus přebírají uměle inteligentní virtuální a robotičtí agenti stále více podpůrné terapeutické intervence, které dříve nabízeli výhradně vyškolení a kvalifikovaní zdravotníci. Aby byla umožněna odpovědná klinická implementace prostředků umělé inteligence pro péči o duševní zdraví, je třeba identifikovat a řešit etické a sociální důsledky rostoucího používání umělé inteligence v této oblasti.

Směřováním vývoje umělé inteligence a analýzou možných etických výzev v oblasti péče o duševní zdraví se zabývá několik odborných studií. Jedna z nich kritizuje současný stav s tím, že stávající objem navrhovaných principů se stal matoucí, což představuje dva potenciální problémy: buďto budou různé soubory etických principů pro umělou inteligenci podobné, což by ale vedlo ke zbytečnému opakování a redundanci, nebo ke zmatku a nejednoznačnosti, pokud se výrazně liší. Studie⁸² Luciana Floridiho, italského filozofa vyučujícího na Oxfordské univerzitě, navrhuje řešení pro sloučení velkého množství principů eticky odpovědné umělé inteligence. Jiná studie⁸³ je zaměřená na hledání principů pro udržení rovnováhy mezi etickými ohledy a transformací zdravotnictví s pomocí umělé inteligence a syntézou relevantních znalostí z řízení umělé inteligence a navrhuje eticky odpovědný řídicí rámec iniciativ pro vývoj umělé inteligence, zahrnující pět hlavních témat pro vývojáře, zdravotnické odborníky a tvůrce politik. Tato témata jsou shrnuta ve zkratce SHIFT - *Sustainability, Human centeredness, Inclusiveness, Fairness a Transparency*. Tato studie je zajímavá tím, že do svých etických

⁸² FLORIDI, Luciano. A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review* 2019.

⁸³ HAYTHAM, Siala. Shifting artificial intelligence to be responsible in healthcare: A systematic review, *Social Science & Medicine* 2022.

úvah včleňuje i jeden z významných etických konceptů – etiku ctností, která je většinou mimo zorné pole současných bioetiků. Jeden typ studie je zaměřený na souhrn zkušeností a očekávání ze strany několika klinických pracovníků, kteří zdůrazňují etické principy nutné pro budoucí širší používání umělé inteligence v poradenství⁸⁴. Poslední typ studie⁸⁵ je zaměřený nejenom na syntézu existujících principů s rozšířením zorného pole jak na vývoj, tak na užití prostředků umělé inteligence, ale i na způsoby, jak vývoj a užívání řídit tak, aby byl výsledkem vývoje a výroby eticky odpovědný produkt.

6.2. Etické výzvy pro budoucí vývoj a užití umělé inteligence

V dalším výkladu se zaměřím na etické zhodnocení budoucích výzev širšího nasazení umělé inteligence při péči o duševní zdraví a zamyslím se nad možnými opatřeními, kterými by mohla být řízena rizika při výskytu potenciálních etických problémů. Vycházím z rešerší odborných textů na dané téma a zavedených principů lékařské etiky, ze kterých byla provedena analýza etických a sociálních aspektů vybraných aplikací umělé inteligence v oblasti péče o duševní zdraví. Aby bylo možné provést komplexní hodnocení, byla analýza strukturována do následujících tří kroků:

- a. posouzení potenciálních přínosů
- b. analýza zastřešujících etických problémů a obav
- c. diskuse o konkrétních etických a sociálních otázkách intervencí.

Virtuální a robotičtí agenti využívající umělou inteligenci jsou stále více k dispozici nejen pro oblast podpory duševního zdraví na relativně nízké úrovni, jako je navození pohodové nálady nebo nenásilné sociální interakce, ale provádějí i některé podpůrné terapeutické intervence, které dříve nabízeli kvalifikovaní zdravotní pracovníci, jako jsou psychoterapeuti a psychiatři. Podstatné je, že takoví virtuální nebo robotičtí pomocní terapeuti v sobě obsahují algoritmus umělé inteligence, jenž reaguje nezávisle na jakémkoli odborném lidském vedení klienta nebo pacienta prostřednictvím virtuálně

⁸⁴ FULMER, Russel. The Ethics of Psychological Artificial Intelligence: Clinical Considerations. *Counselling and Values* 2021.

⁸⁵ FISKE, Amelie. A Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy, *J Med Internet Res* 2019.

ztělesněné přítomnosti, jako je ikona obličeje, nebo fyzicky ztělesněné přítomnosti, jako je robotické rozhraní modelující v základních rysech lidské tělo a obličej. Jedním z nastupujících prvků umělé inteligence používané v oblasti péče o duševní zdraví a duševní pohodu, který nemusí být z etické perspektivy bezproblémový, je použití vizualizace lidského těla a tváře. Jako takové se tyto nově vznikající aplikace liší od mnoha druhů webové terapie, která obvykle zahrnuje buď lidského terapeuta v online terapii na dálku nebo samotného pacienta, který nezávisle na terapeutovi pracuje s manuály, dotazníky nebo jinými svépomocnými materiály.

Začleněné aplikace umělé inteligence v péči o duševní zdraví nesou určitou naději na zlepšení kvality péče a optimalizaci nákladů, protože výdaje spojené s tímto typem podpůrné terapie jsou nižší než u tradiční péče. Kromě toho také slibují oslovit populaci s nedostatečnými službami, jež potřebuje služby pečující o duševního zdraví, a zlepšit životní možnosti pro zranitelné skupiny populace. Existuje však přetrvávající rozpor mezi současným rychlým rozvojem umělé inteligence a její aplikací pro podporu duševního zdraví a mnohdy oprávněnou neochotou přijímat tyto nástroje do klinického prostředí ze strany kliniků, ale i pacientů. Jedním z důvodů bude patrně to, že ačkoli studie často zkoumají účinnost nebo etické využití jedné aplikace, zřídka se zvažují důsledky pro integraci umělé inteligence v oblasti duševního zdraví v širším měřítku a zcela opomíjejí zkoumání dlouhodobosti terapeutických účinků.

Je patrné, že virtuálně a fyzicky ztělesněné prostředky umělé inteligence – roboty – a aplikace by mohly mít významný potenciál v péči o duševní zdraví. Společenské a etické důsledky jejich aplikace však vyžadují další zkoumání, aby bylo možné identifikovat relevantní obavy týkající se důvěry, soukromí a autonomie, jakož i předvídat obavy, které se mohou objevit v budoucnu. Identifikace širších etických a společenských důsledků při užití těchto prostředků umělé inteligence je zásadní pro jednání o nejlepších výzkumných a lékařských postupech v inovativní péči o duševní zdraví ve smyslu *Lege artis* mezi vývojáři systému a regulátory zdravotních politik. Budoucí výzkum se bude odehrávat v konkrétních oblastech, jako jsou virtuální roboti nebo antropomorfičtí roboti pečující o duševní zdraví, přičemž ke každé oblasti je potřeba přistupovat se specifickou etickou opatrností.

Výše popsané oblasti s antropomorfičtými robotickými zařízeními a aplikacemi virtuálních robotů musí být ještě integrovány do širokého klinického použití. S ohledem

na rychlost výzkumných a vývojových trajektorií těchto aplikací je však rozumné očekávat, že terapeutičtí chatboti, avataři, sociálně asistenční zařízení se již brzy skutečně promítnou do širších klinických aplikací. V některých případech jsou již k dispozici počáteční etická hodnocení, většina studií se však zaměřuje na jedinou aplikaci, jíž je péče o seniory. Obecně platí, že pro většinu aplikací dosud nebyly provedeny rozsáhlé rigorózní výzkumné studie nebo - jak je tomu u některých konverzačních robotů - jsou tyto studie stále v pilotních fázích. Existují i rozsáhlejší expertní hodnocení některých chatbotů⁸⁶.

Problematické je rovněž přijímání aplikací ze strany pacienta. I ve webových aplikacích bez umělé inteligence jsou důkazy o přijetí pacientem a výsledcích léčby v běžné péči stále omezené a nejednoznačné. Na rozdíl od chatbotů, analyzovaných v předchozích kapitolách, bylo dosud provedeno velmi málo výzkumu o tom, jak jsou pacienti přijímáni antropomorfičtí roboti a jaké jsou případné výsledky léčby v oblastech duševního zdraví. Jako u každé lékařské inovace lze účinky, dopady a klinickou užitečnost aplikací plně posoudit až poté, co bude k dispozici více testů. Aby byla umožněna eticky odpovědná a citlivá inovace a klinické použití v oblasti duševního zdraví, je nezbytná další a hlubší analýza etických a sociálních důsledků antropomorfních robotů užívajících umělou inteligenci. Včasná identifikace etických problémů může výzkumníkům, návrhářům a vývojářům pomoci zvážit tyto obavy při návrhu a konstrukci další generace aplikací pro duševní zdraví.

Všechny výše uvedené aplikace využívající umělou inteligenci slibují významné přínosy pro oblast duševního zdraví a tím splňují mnoho aspektů etického principu Beneficence⁸⁷. Z klinického hlediska má použití aplikací umělé inteligence potenciál otevřít nové cesty pro intervenci v místech, kde jsou významně neuspokojené zdravotní potřeby, a to většinou z důvodu malé dostupnosti psychoterapeutů. Intervence aplikací umělé inteligence mohou být zvláště vhodné pro včasné odhalování obav o duševní

⁸⁶ Woebot: A Professional Review (30.8.2019) [2022-5-31]. < <https://onemindpsyberguide.org/expert-review/woebot-an-expert-review/>>.

⁸⁷ PŘÍHODA, Petr. Etika : Čtvero základních principů lékařské etiky [online]. [cit. 2011-12-23]. <<http://www.lf2.cuni.cz/ustav-lekarske-etiky-a-humanitnich-zakladu-mediciny-2-lf-uk/etika>>.

zdraví, pro oslovení vysoce rizikových skupin, jako jsou lidé s posttraumatickým syndromem, nebo pro ty, kteří se obávají sociálního stigmatu spojeného s psychoterapií.

V některých případech mohou pacienti reagovat pozitivně a produktivně na skutečnost, že protějšek není lidský terapeut, například když se jedná o pacienty se sociální fobií nebo generalizovanou úzkostnou poruchou. Ve službách péče o duševní zdraví tedy mohou být pro některé pacienty vhodnější virtuální nebo robotické aplikace, které snižují rozpaky pacienta, když žádají o konkrétní informace nebo služby, nebo když by takový pacient měl pocity studu při přiznání nedodržování léčebného plánu stanoveného klinikem. Antropomorfní roboty s umělou inteligencí určené pro péči o duševní zdraví by také mohly pomoci posílit postavení konkrétních skupin pacientů, kteří jsou méně obeznámeni s lékařským systémem, a tím pomoci zlepšit důvěru a otevřenost mezi pacienty a lékařským systémem.

Další důležitou výhodou aplikací umělé inteligence je to, že mnohé z nich jsou nízkoprahové a samoobslužné, takže lidé, co se nenalézají v akutním stavu, mohou volit služby, aniž by museli projít časově náročným procesem screeningu a přijetí do systému zdravotní péče. Pravděpodobně největší přínos aplikací umělé inteligence bude strukturální, kde konkrétně naplní svůj potenciál oslovit populace, které je obtížné léčit tradičními způsoby. Poskytování některých služeb duševního zdraví, například prostřednictvím nízkoprahových, pohodlných podpůrných terapeutických intervencí prostřednictvím chatbotů nebo avatarů, může být zvláště přínosné pro populace žijící v prostředí s nedostatkem zdrojů. Pro ty, kteří žijí v odlehlých nebo venkovských oblastech nebo v prostředí, kde jsou místní služby duševního zdraví málo dostupné, mohou aplikace umělé inteligence zvýšit dostupnost a poskytnout některé alespoň minimální služby péče o duševní zdraví tam, kde jinak chybí. Kromě toho je pravděpodobné, že existují jedinci, co z různých důvodů nereagují na tradičnější klinické služby a preferují nízkoprahové intervence, které lze provádět v soukromí jejich domova nebo na cestách. U všech těchto pacientů by aplikace umělé inteligence mohly doplňovat stávající služby nebo představovat vstupní bod pro provádění standardnějších klinických intervencí v budoucnu.

Stručně řečeno, intervence prostředky umělé inteligence mohou nabídnout zcela nové způsoby léčby, které jsou potenciálně úspěšnější než tradiční způsoby, ať už proto, že

oslovují těžko dosažitelné populace, nebo proto, že na ně pacienti lépe reagují. Zda a za jakých podmínek tomu tak je, vyžaduje další šetření.

Pokud by byly aplikace s umělou inteligencí integrovány do škálovaného poskytování služeb, mohly by poskytovat podporu pro mírné případy deprese a jiné neakutní stavy, a tak pomoci zdravotníkům věnovat více času těžším případům. S ohledem na celkově rostoucí zátěž nemocí v oblasti duševního zdraví, jako jsou například úzkostné poruchy, a na pozadí omezených zdrojů je třeba zvážit tyto důležité výhody. Je však pravděpodobné, že aplikace prostředků umělé inteligence nemusí být pozitivně přijímána všemi odborníky na péči o duševní zdraví a někteří mohou mít dokonce vážné pochybnosti o jejím použití z důvodu etických nebo klinických obav. Je skutečností, že dosud nebyl proveden žádný zásadní přezkum přijímání umělé inteligence ve specifických oblastech péče duševního zdraví, a tato oblast proto vyžaduje další výzkum.

Jedna z nejzávažnějších výzev se nalézají v oblasti prevence negativních následků a etiky spojená s pořizováním, zpracováváním a uchováváním dat. Aby bylo vyhověno všeobecně přijímanému etickému principu nonmaleficence⁸⁸, je zapotřebí důkladnější výzkum aplikací umělé inteligence v oblasti péče o duševní zdraví. Tím by se předešlo poškození jak při terapeutických sezeních, tak v případech, kdy by roboty mohly selhat nebo pracovat nepředvídatelným nežádoucím způsobem. Přestat fungovat nebo selhat by mohli i chatboti a avataři. Proto je třeba prodiskutovat, zda by na zařízení umělé inteligence včetně virtuálních systémů a volně dostupných aplikací pro duševní zdraví měl být vyžadován stejný druh přísného hodnocení rizik a regulačního dohledu, jakému podléhají i jiná zdravotnická zařízení před schválením pro klinické použití.

Podobně jako u jiných systémů používaných v lékařských zařízeních, vyžaduje použití jakýchkoli aplikací umělé inteligence v péči o duševní zdraví pečlivé zvážení formě zabezpečení dat zařízení, kterému jsou předávána osobní zdravotní informace, a způsobů, jakými jsou zadaná data používána. V otázkách důvěrnosti, soukromí informací a bezpečné správy dat shromážděných inteligentními virtuálními agenty a asistenčními

⁸⁸ PŘÍHODA, Petr. Etika : Čtvero základních principů lékařské etiky [online]. [cit. 2011-12-23]. <<http://www.lf2.cuni.cz/ustav-lekarske-etiky-a-humanitnich-zakladu-mediciny-2-lf-uk/etika>>.

roboty jsou zapotřebí jasné standardy, stejně jako v jejich použití pro sledování návyků, pohybu a dalších interakcí mezi pacientem a jeho okolím. Obavy týkající se soukromí mohou být zesíleny, protože množství shromážděných údajů se neustále rozšiřuje. Lze například předpokládat, že aplikace, které integrují video data a mohou například detekovat sebevražedné symptomy, by musely mít specifickou ochranu soukromí pro sdělování citlivých informací či informací týkajících se jiných osob, než jen pouhý souhlas pacienta.

Dalším zřetelným rizikem je nedostatek standardů pro vývoj, klinickou integraci a školení odborného personálu. Vzhledem k tomu, že umělá inteligence je jednou z nejnovějších a nejrychleji se měnících oblastí psychologického a psychiatrického výzkumu a léčby, tak nejsou stávající právní a etické rámce často s těmito změnami úžeji sladěny. Existuje reálné riziko, že nepokryté oblasti dané nesladěním aplikačního a etického rámce budou vyřešeny až poté, co dojde k naplnění rizika. To bývá případ i v jiných lékařských technologiích; vzhledem k rychlému tempu převádění umělé inteligence do praxe v prostředích, kde tradiční systémy hodnocení zdravotnických technologií a lékařského dohledu nejsou plně použitelné.

Ačkoli je obtížné předvídat etické a právní otázky, které se objeví spolu s budoucím vývojem, je nutné aktivně uvažovat o vhodnosti regulovat umělou inteligenci pro péči o duševní zdraví. Objevují se iniciativy pro vytvoření metodických pokynů, které již byly zmíněny v předchozí kapitole, jako je příklad Etického rámce pro dobrou společnost umělé inteligence nebo iniciativa Evropské unie a jejího expertního týmu. Problém ale je, že doposud neexistují žádné instrukce a standardy, specifické pro vývoj prostředků umělé inteligence v oblasti péče o duševního zdraví. Kromě nedostatku instrukcí pro vývoj těchto intervencí pro otázky designu a použití nejsou k dispozici ani žádné rámce, jak by se zdravotníci mohli efektivně zapojit a zaučovat na předpokládané zvýšené klinické používání prostředků umělé inteligence. Znamená to, že ačkoli existuje stále větší množství akademické i populární literatury o tom, jak lze umělou inteligenci začlenit do klinické praxe v oblasti duševního zdraví, tak stále chybí odborné poradenství na dostatečně vysoké úrovni ohledně optimálního využití umělé inteligence v oblasti péče o duševní zdraví. Proto jsou zapotřebí další etické pokyny specifické pro pomoc odborníkům v oblasti duševního zdraví, kteří budou dohlížet na pacienty, co mají nebo možná budou využívat prostředky umělé inteligence při péči o svoje duševní zdraví.

Další hrozba spočívá v potenciálním zneužití umělé inteligence ke snížení poskytování stávajících služeb. Eticky zodpovědná integrace prostředků umělé inteligence do stávajících služeb by měla také zvážit otázky spravedlivého poskytování péče o duševní zdraví. Existuje obava, že začlenění prostředků umělé inteligence do péče o duševního zdraví by mohlo ospravedlnit nahrazení existujících služeb, což by mělo za následek rostoucí počet méně dostupných reálných zdravotních zdrojů, čímž by se potenciálně prohloubily stávající nerovnosti v oblasti zdraví.

Mnoho zastánců důkladnější regulace oprávněně trvá na tom, že ačkoli jsou chatboty koncipovány na psychoterapeutických přístupech založených na důkazech⁸⁹, neměly by nahrazovat psychoterapeuty. V některých případech by mohla vyhovovat forma smíšené péče zahrnující jak osobní, tak virtuální formy terapie. Modely smíšené péče potenciálně nabízejí příležitost čerpat ze silných stránek aplikací umělé inteligence i osobního klinického dohledu. To, zda je či není vhodné implementovat aplikace umělé inteligence do péče o duševní zdraví, však částečně závisí na dostupnosti dalších zdrojů v této oblasti. Jak bylo uvedeno, v případech s omezenými službami v oblasti duševního zdraví by aplikace umělé inteligence mohly poskytnout potřebný zdroj, který je rozhodně lepší než služby žádné. V tuto chvíli však služby duševního zdraví se zapojením umělé inteligence nenahrazují ani nezastupují druh robustní, víceúrovňové péče o duševní zdraví dostupné v systémech zdravotní péče s vysoce kvalifikovanými zdroji, jako je tomu i v Česku. Vhodné zvážení status quo zdrojů duševního zdraví v každém kontextu je tedy z etického principu spravedlnosti vysoce relevantní.

S ohledem na aplikaci nástrojů umělé inteligence v péči o duševní zdraví je třeba mít na paměti řadu specifických problémů. Odborníci v oblasti duševního zdraví mají etickou odpovědnost za informování ostatních poskytovatelů služeb stejně jako třetí strany nebo úřady, pokud z komunikace s pacientem vyplývá, že je hrozbou pro sebe nebo pro jinou osobu. Je například nutné určit, jak by to fungovalo u intervencí s umělou inteligencí, zejména pokud neexistuje žádný dohled kvalifikovaného zdravotnického pracovníka nad interakcí mezi aplikací umělé inteligence a pacientem. Není jasné, kdy a jak by volně

⁸⁹ Viz kapitola o léčbě úzkostných stavů

dostupní virtuální agenti a chatboti, spolehlivě propojili ohroženého jedince s vhodnými záchrannými službami, včetně hospitalizace a další případné ochrany. Co by se však mělo dělat, pokud například terapeutický chatbot prostřednictvím řečových vzorů zjistí, že jedinec je vystaven vyššímu riziku sebepoškození, a přesto v dané oblasti nejsou dostupné vhodné služby, které by mu zajistil?

Aplikace umělé inteligence zapojené do terapeutických vztahů s klienty budou pravděpodobně také muset být vázány podobnými etickými pravidly jako ty, které zavazují odborníky v oblasti duševního zdraví. Dosud však není zcela jasné, jaká by měla být a vůči komu uplatňována povinnost tohoto druhu péče o prostředky s umělou inteligencí. Určitou možností by bylo vždy nařídít supervizi prostřednictvím kvalifikovaného klinika duševního zdraví. V případě, kdy reálný terapeut hodnotí pacientovo vyjádření k sebepoškození, měl by brát také v úvahu kontextové informace ve své interpretaci úrovně rizika. Zda a do jaké míry jsou toho robotičtí terapeuti schopni, zůstává nejasné. Mnoho aplikací umělé inteligence je však dostupných mimo zavedená prostředí duševního zdraví. Některé z nich jsou například dostupné v archivu Google Play nebo Microsoft Store, zvyšuje se i kapacita počítačových metod k identifikaci a predikci psychiatrických onemocnění, stejně jako jejich schopnosti terapeutické interakce a komunikace. Předmětem dalších debat proto zůstává, zda by měl být vždy zajištěn dohled nad aplikacemi umělé inteligence v oblasti péče o psychické zdraví a jak by mohl být takový požadavek úspěšně implementován.

Významná je i výzva k respektování a ochraně autonomie pacienta. Požadavek je, aby aplikace umělé inteligence v péči duševního zdraví umožnily respektování autonomie pacienta⁹⁰. Aplikace umělé inteligence je nová technologie, které by měla zaručit, že pacienti na určité úrovni chápou, jak aplikace, chatbot nebo avatar fungují, aby se zajistilo, že si pacient nezamění systém s umělou inteligencí za aplikaci řízenou člověkem. Například by bylo problematické, kdyby pacient předpokládal, že na druhé komunikující straně chatbota je i lékař nebo terapeut, jenž komunikuje anebo pouze kontroluje zasílané zprávy.

⁹⁰ PŘÍHODA, Petr. Etika : Čtvero základních principů lékařské etiky [online]. [cit. 2011-12-23]. <<http://www.lf2.cuni.cz/ustav-lekarske-etiky-a-humanitnich-zakladu-mediciny-2-lf-uk/etika>>.

Bude třeba vyřešit to, jak by měly aplikace umělé inteligence vyhodnotit, zda pacient plně porozuměl poskytnutým informacím při udělování souhlasu, a jak postupovat v případech, kdy není možné souhlas poskytnout jednotlivcům, jako jsou děti, pacienti s demencí, mentálně postižení nebo těm, kteří jsou v akutních fázích schizofrenie. Další oblastí zvláštního zájmu ve vztahu k otázkám podpory autonomie při používání umělé inteligence v péči o duševní zdraví je zapojení zranitelných skupin obyvatelstva. Ukázalo se, že tito lidé jsou poslušnější, když je robot požádá, aby něco udělali, než když je o to požádá člověk⁹¹. Ačkoli by to mohlo vést k lepším výsledkům při pomoci pacientům s autismem nebo těm, kteří potřebují provést obtížné změny chování, existuje obava, že by lidé mohli být manipulováni nebo donuceni dělat věci, které by neměli dělat nebo které si plně nepromysleli, ať už kvůli novosti zařízení nebo z důvodu nedostatku blízkých lidí, s nimiž lze diskutovat o různých možnostech.

Důležitá je i výzva, zda je terapie poskytovaná mimolidským prostředkem umělé inteligence ještě vůbec terapií. Obecně zůstává otázkou, zda existují aspekty terapeutického setkání, jichž nelze prostřednictvím umělé inteligence dosáhnout. Některé terapeutické přínosy může být obtížné předvídat nebo mohou být vysoce specifické pro vztah konkrétního jedince k jeho terapeutovi. Vzhledem k tomu, že systémy umělé inteligence stírají dříve předpokládané hranice mezi realitou a fikcí, mohlo by to mít na pacienty komplexní dopad. Podobně jako u terapeutických vztahů existuje riziko přenosu emocí, myšlenek a pocitů z pacienta na robota. Vzhledem k tomu, že mnoho cílových populací je zranitelných kvůli své nemoci, věku nebo životní situaci ve zdravotnickém zařízení, existuje další obava, že pacienti by byli zranitelní při práci s robotem kvůli své touze po společnosti nebo kvůli potřebě cítit, že je o ně postaráno. Na rozdíl od terapeuta však na druhé straně tohoto spojení není žádná živá osoba. Zda budou robotičtí terapeuti někdy schopni se s takovým přenosem adekvátně vypořádat, není v současné době zřejmé.

Techničtějšího rázu může být další výzva související s etickými problémy v algoritmech, které umělá inteligence používá. Je prokázáno, že existující lidské

⁹¹ BROADBENT, Elisabeth. Interactions with robots: the truths we reveal about ourselves. *Annu Rev Psychol* 2017. s. 627-652.

předsudky mohou být skrze designery umělé inteligence promítnuty do algoritmů, čímž se potenciálně mohou posílit existující formy sociální nerovnosti. To vyvolává obavy, že zařízení pro duševní zdraví s umělou inteligencí by také mohla mít v sobě předsudky, které mají potenciál vyloučit nebo poškodit pacienta nezamýšleným způsobem. Příkladem mohou být předsudky sexistické nebo rasistické povahy nebo necitlivé kulturní odsudky. V návaznosti na výzvy k transparentnosti algoritmů, by mohly být algoritmy používané v aplikacích s umělou inteligencí pro účely duševního zdraví podobně přístupné nezávislé kontrole.

Výzvy se také týkají obavy z dlouhodobých účinků. Bylo například zaznamenáno, že dlouhodobé používání intervencí umělé inteligence by mohlo vést k tomu, že někteří pacienti nebo skupiny pacientů budou k těmto aplikacím příliš připoutáni. Nedávná studie zdůrazňuje, že roboti, jejichž cílem je zmírnit osamělost nebo poskytnout emocionální pohodu, představují riziko, že se na nich pacienti, se kterými pracují, mohou stát patologicky závislími⁹², podobně jako je tomu u lidí. Vliv inteligentních robotů na vztahy, a to jak mezi lidmi a roboty, tak mezi lidmi a lidmi, je oblastí, která vyžaduje další zkoumání, stejně jako potenciální vliv na identitu, jednání a sebevědomí u jednotlivých pacientů. Existuje riziko, že pokud by intervence robota nebylo možné využít na zlepšení interakce pacienta s ostatními lidmi, tak zůstanou pouze způsobem zlepšení vztahu pacienta se strojem, nebo v horším případě situací, v níž se dále omezují vztahy mezi pacientem a ostatními lidmi.

Integrace zařízení umělé inteligence do našich každodenních životů nepochybně mění společenská očekávání a způsoby komunikace. Mezi komunikací se zařízením umělé inteligence a komunikací s jiným člověkem jsou však zásadní rozdíly. Existující zjištění naznačují, že někteří uživatelé často mluví s virtuálními pomocníky v mobilních telefonech, jako je Siri nebo Alexa, ostřeji nebo hruběji než s člověkem. Důležité je, že vnímání zařízení se může u uživatelů lišit. Děti často chápou tato zařízení jinak než dospělí, někdy zařízení připisují lidské vlastnosti nebo věří, že zařízení má uvnitř živého člověka. Extrapolací z tohoto příkladu je jasné, že způsoby interakce jednotlivců s

⁹² CRESSWELL, Kathrin. Health care robotics: qualitative exploration of key challenges and future directions. *J Med Internet Res* 2018.

aplikacemi umělé inteligence mohou mít vážné důsledky pro komunikaci a sociální interakci. Jak se to bude vyvíjet, až bude mít více pacientů příležitost interagovat s aplikacemi umělé inteligence v rámci péče o duševní zdraví, vyžaduje proto další empirické zkoumání, aby bylo možné včas zachytit problematické trendy a přijmout nápravná opatření pro eticky odpovědný budoucí vývoj. Navíc je pravděpodobné, že jak se budou vyvíjet inteligentnější a autonomnější zařízení umělé inteligence, budou lidské vztahy s nimi ještě komplikovanější.

Ve světle prokázaných přínosů a potenciálu, jako je rozšíření dosahu služeb na populaci s nedostatečnými službami nebo zlepšení stávajících služeb poskytovaných odborníky na duševní zdraví, se umělá inteligence ukázala jako možný slibný přístup v specifických oblastech péče o duševní zdraví. V současnosti je však kvalita výzkumu využití prostředků umělé inteligence v psychiatrii, psychologii a psychoterapii různorodá a existuje výrazná potřeba robustnějších studií, včetně randomizovaných studií o výhodách a potenciálních škodách současných a budoucích aplikací. V umělé inteligenci jsou stále nově vznikající oblasti a jakákoliv analýza etických důsledků může být z tohoto pohledu pouze předběžná.

Na základě úvah uvedených v tomto dokumentu je však opodstatněných několik závěrů a doporučení: Je nutné vypracovat jasné závěry a z nich vyplývající pokyny, zda vůbec a které aplikace umělé inteligence by měly podléhat standardnímu hodnocení zdravotnických technologií a měly by vyžadovat schválení regulačními orgány. To by mělo zahrnovat soubor širších ustanovení pro používání služeb umělé inteligence mimo dohled zdravotnického pracovníka. Profesionální sdružení v oblasti duševního zdraví by měla vypracovat instrukce pro nejlepší využití umělé inteligence ve službách péče o duševní zdraví a také doporučit, jak školit a připravovat psychology a psychiatry na budoucí masovější používání prostředků umělé inteligence v oblasti duševního zdraví, včetně modelů smíšené péče. Nástroje umělé inteligence v oblasti duševního zdraví by měly být považovány za doplňkový zdroj v podpůrných službách péče o duševní zdraví. Neměly by být používány jako záminka pro omezení poskytování vysoce kvalitní péče vyškolenými odborníky v oblasti duševního zdraví a bude nutné posoudit jejich vliv na dostupnost a využívání stávajících služeb péče o duševní zdraví.

U všech aplikací nabízených mimo poskytované péče o duševní zdraví, by se mělo vyžadovat, aby prokázaly spolehlivé způsoby hodnocení rizik a měly by obsahovat

doporučení k příslušným službám. Je třeba vypracovat návod, jak implementovat aplikace způsobem, který respektuje autonomii pacienta, například pokud jde o to, kdy a jak je vyžadován souhlas a jak nejlépe řešit záležitosti zranitelnosti, manipulace, nátlaku a soukromí.

Algoritmy umělé inteligence v oblasti duševního zdraví je třeba prozkoumat, například kvůli zkreslení. V ideálním případě by zdravotníci měli být vyškoleni v tom, jak svým pacientům sdělovat roli algoritmů používaných v různých aplikacích, které mohou používat nebo zvažují použít. Takové algoritmy by pak měly být otevřené pro veřejnou diskusi a formování.

Zvýšené používání umělé inteligence by mělo být doprovázeno výzkumem, jenž zkoumá přímé i nepřímé účinky na terapeutický vztah, jiné vztahy mezi lidmi a účinky na individuální sebevědomí, jednání a identitu. Je třeba monitorovat dlouhodobé účinky, od zdravotního redukcionismu po zvýšenou objektivizaci a dopady na naše chápání toho, co znamená být člověkem obklopeným prostředím zčásti ovládaném umělou inteligencí.

V návaznosti na studii⁹³ a na základě předchozích úvah je možné učinit pro oblast řízení etických rizik při dalším vývoji umělé inteligence pro zdravotnictví několik souhrnných závěrů a vydat následující doporučení:

- Je nutné vypracovat jasné pokyny, zda a které aplikace umělé inteligence by měly podléhat standardnímu hodnocení zdravotnických technologií a vyžadovat schválení regulačními orgány.
- Profesní sdružení v oblasti duševního zdraví by měly vypracovat pokyny pro optimální využití umělé inteligence ve službách péče o duševní zdraví a také doporučení, jak školit a připravovat lékaře a psychology na budoucí rozšířené používání umělé inteligence v oblasti duševního zdraví, včetně modelů smíšené péče.
- Nástroje umělé inteligence v oblasti duševního zdraví by měly být považovány za doplňkový zdroj ve službách podpůrné péče o duševní zdraví a měly by být užívány pod supervizí příslušného zdravotníka.

⁹³ FISKE, Amelie. A Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy, *J Med Internet Res* 2019.

- Umělá inteligence by měla být používána transparentně. Je třeba vypracovat návod, jak implementovat aplikace způsobem, který respektuje autonomii pacienta.
- Algoritmy umělé inteligence v oblasti duševního zdraví je třeba prozkoumat, například kvůli zkreslení. V ideálním případě by zdravotníci měli být vyškoleni v tom, jak svým pacientům sdělovat roli algoritmů používaných v různých aplikacích.
- Zvýšené používání umělé inteligence by mělo být doprovázeno výzkumem, jenž zkoumá přímé i nepřímé účinky na terapeutický vztah, jiné vztahy mezi lidmi a účinky na individuální sebevědomí, jednání a identitu.
- Je třeba monitorovat dlouhodobé účinky, od zdravotního redukcionismu po zvýšenou objektivizaci a dopady na naše chápání toho, co znamená být člověkem v prostředí, které zčásti řídí umělá inteligence.

Závěr

Tato diplomová práce se věnovala etickým otázkám spojeným s použitím vybraných prostředků umělé inteligence pro léčbu a zvládnání úzkostných stavů. Zaměřila se na napětí mezi podpůrnou terapií poskytovanou prostředky umělé inteligence a psychoterapií prováděnou kvalifikovanými psychoterapeuty. V první části jsem načrtl terminologické pole zahrnující oblast umělé inteligence a jejich nástrojů, léčbu úzkostí, etiku psychoterapie a další podstatné pojmy, které v rámci řešené problematiky používám a jejichž vymezení slouží k plnému pochopení výkladu. Dále jsem představil základní rysy kognitivně behaviorální terapie, jejíž model léčby úzkostných stavů a úzkostných poruch je replikován do algoritmů využívaných v prostředcích umělé inteligence určených pro léčbu a podporu zvládnání úzkostných stavů. Představil jsem také základní informace o regulatorních rámcích psychoterapie, vytvořených na základě základních principů bioetiky a rozšířené o princip související se specificitou psychoterapie.

V druhé části jsem přiblížil metodu, která byla použita pro zkoumání dané problematiky.

Ve třetí části jsem popsal různé koncepty etických principů používaných v psychoterapii a zvláštní podkapitulu jsem věnoval pohledu na etiku psychoterapie z perspektivy bioetiky a specificky retrospektivní bioetiky tak, jak jí navrhuje český etik a filosof Josef Kuře. Další podkapitulu jsem věnoval srovnání různých konceptů etiky v oblasti umělé inteligence. Její úvod jsem věnoval vysvětlení významu Turingova testu pro rozvoj etiky umělé inteligence používané pro péči o duševní zdraví a velký prostor jsem věnoval vysvětlení pojetí etiky umělé inteligence, zvolené zeměmi Evropské unie. Hlavním důvodem pro tuto volbu byl fakt, že politiky Evropské unie z této oblasti se budou dříve nebo později transponovat do práva České republiky a stanou se tak aktuální i pro úvahy o vývoji a systematictější nasazení umělé inteligence v oblasti péče o duševní zdraví v Česku.

Ve čtvrté části jsem se zaměřil na výklad pojetí léčby úzkostných poruch a léčebných postupů pro pomoc při zvládnání úzkostných stavů. Zvláštní pozornost jsem věnoval výkladu specifického psychotherapeutického směru v kognitivně behaviorální terapii, jehož model léčebného působení byl zčásti převzat do algoritmů prostředků umělé inteligence zaměřené na léčbu úzkostných poruch.

Na základě teoretických východisek formulovaných v prvních čtyřech částech jsem v páté části provedl etickou reflexi vybraných konkrétních konverzačních nástrojů umělé inteligence používaných v terapii úzkostných a depresivních stavů. V první podkapitole jsem reflektoval podstatu konverzačních prostředků umělé inteligence včetně úvodních etických úvah o jejich použití, ve druhé podkapitole jsem uvedl základní charakteristiky a hodnocení vybraných chatbotů-konverzačních prostředků umělé inteligence určené pro podporu při zvládání a léčbě úzkostných stavů. Ve třetí podkapitole jsem odpověděl na hlavní výzkumnou otázku tím, že jsem podrobil vybrané chatboty etické reflexi. Specifickou pozornost jsem věnoval čtvrté podkapitole, kde jsem řešil etické limity použití chatbotů včetně etické problematiky vztahu mezi pacientem a chatbotem. V závěrečné páté podkapitole jsem použil etická východiska terapie využívající dialogický přístup vycházející z dialogického personalismu k analýze limitů použití chatbotů v terapii.

V šesté závěrečné části jsem se prospektivně zaměřil na další rozvoj nasazování umělé inteligence v oblasti péče o duševní zdraví a odpověděl jsem na vedlejší výzkumnou otázku zkoumáním možných etických implikací tohoto předpokládaného vývoje. Naznačil jsem rovněž směr a potřebnost dalších kroků nutných pro systematické budování eticky odpovědné umělé inteligence v oblasti péče o duševní zdraví.

Z práce je zřejmé, že problematika umělé inteligence v péči o duševní zdraví je multioborová, komplexní a obsahuje nejen etické, psychologické, psychiatrické a technické kontexty, ale i otázky právní, společenské a organizačně zdravotnické. Přestože v současnosti probíhají v mnoha státech práce na tvorbě politik tvorby a provozování eticky únosné umělé inteligence, jak se bylo možné dočíst ve třetí kapitole, je vývoj samotné umělé inteligence podstatně dynamičtější. Tato rozdílná dynamika přináší řadu závažných rizik, která vedou konzervativnější kliniky ke spíše obezřetnému přístupu. S tímto přístupem předběžné opatrnosti se osobně plně ztotožňuji, protože se v praxi setkávám s případy, kdy by použití chatbotů mohlo vést k závažnému porušení etických principů, a to zejména maleficience a autonomie.

Z etického hlediska patří k důležitým přínosům aplikací umělé inteligence v oblasti péče o duševního zdraví nové způsoby podpůrné léčby, příležitosti poskytování služby i pro obtížně dosažitelnou populaci pacientů, lepší recepce pacientů z důvodu eliminace stigmatizace a uvolnění času pro psychoterapeuty a psychiatry. Mezi zastřešující etické

problémy a zájmy byly identifikovány různé otázky etiky dat, nedostatek vodítek pro eticky odpovědný vývoj aplikací umělé inteligence, jejich klinickou integraci a školení zdravotnických odborníků; nesoulady v etických a regulačních rámcích umělé inteligence; potenciál pro zneužití, včetně použití technologií k nahrazení zavedených služeb, čímž by se potenciálně naopak prohloubila stávající nerovnosti v oblasti zdraví. Ze specifických výzev mě osobně zaujala nutnost spojená s budoucím předpokládaným zvýšeným používáním umělé inteligence v oblasti psychoterapie, kdy je potřeba organizovat výzkum zkoumající z dlouhodobé perspektivy přímé i nepřímé účinky na terapeutický vztah, na vztahy mezi lidmi vůbec a účinky na individuální sebevědomí, odpovědnost při jednání a identitu uživatele prostředků umělé inteligence.

Během vypracování diplomové práce jsem se přiklonil ke dvěma hlavním závěrům. *První hlavní závěr je, že z etické perspektivy není správné, aby byly prostředky umělé inteligence používány ať vědomě či nevědomě jako alternativa k léčbě duševních poruch prováděné psychoterapeutem, protože by tím byly významně porušeny etické principy autenticity, beneficence a maleficence.* Jejich použití je však možné ve specifických případech pod supervizí pro práci s prostředky umělé inteligence kvalifikovaného psychoterapeuta, a to zejména jako specifická forma podpůrné léčby. *Druhý hlavní závěr je, že umělá inteligence je sice slibným přístupem napříč různými oblastmi podpůrné péče o duševního zdraví, ale je však zapotřebí dalšího výzkumu, který by identifikoval a řešil širší etické a dlouhodobé společenské důsledky využívání těchto technologií, aby bylo možné vyvinout a zavést optimální výzkumné, vývojové a lékařské politiky, postupy a vodítka v inovativní péči o duševní zdraví.*

Seznam použitých zkratk

EFPA	Evropská federace psychologických asociací
APA	Americká psychologická asociace
OECD	Organizace pro hospodářskou spolupráci a rozvoj
AI HLEG	Skupina expertů pro umělou inteligenci při EU
AI	Umělá inteligence
EBT	Terapie založená na důkazech
ACT	Terapie přijetí a závazku

Seznam použité literatury a pramenů

Bibliografické zdroje

DE SAINT LAURENT, Constance. In Defence of Machine Learning: Debunking the Myths of Artificial Intelligence. *Eur J Psychol* 2018. roč.14, č.4. s.734-747.

BLEASE Charlotte. Artificial Intelligence and the Future of Primary Care: Exploratory Qualitative Study of UK General Practitioners' Views. *J Med Internet Res*. 2019 Mar 20;21(3):e12802. doi: 10.2196/12802. PMID: 30892270; PMCID: PMC6446158.

BROADBENT, Elisabeth. Interactions with robots: the truths we reveal about ourselves. *Annu Rev Psychol* 2017. s. 627-652.

BURUK, Banu. A critical perspective on guidelines for responsible and trustworthy artificial intelligence. *Med Health Care and Philos* 2020. roč. 4 čís. 23, s. 387–399.

CLARK, David A.- BECK, Aaron T. *Cognitive therapy of anxiety disorders: Science and practice*. New York: Guilford Press, 2011.

COLLINS, Emily. Drawing parallels in human-other interactions: a trans-disciplinary approach to developing human-robot interaction methodologies. *Philos Trans R Soc Lond B Biol Sci*. 2019 Apr 29;374(1771):20180433. doi: 10.1098/rstb.2018.0433. PMID: 30853002; PMCID: PMC6452243.

CONLIN Williem. Ethical considerations for addressing distorted beliefs in psychotherapy. *Psychotherapy* 2019, roč. 56, č. 4, s. 449-458.

CRESSWELL, Kathrin. Health care robotics: qualitative exploration of key challenges and future directions. *J Med Internet Res* 2018.

ELIZA - A Computer Program for the Study of Natural Language Communication between Man and Machine, *Communications of the Association for Computing Machinery* 1966. s. 36-45.

European Federation of Professional Psychologist Association. *Meta-code of ethics*. Brussels: EFPA, 2005.

FISKE, Amelie. A Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy, *J Med Internet Res* 2019.

FITZPATRICK, Kartleen. 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*, 2017 roč. 4, č. 2.

GIBNEY, Elisabeth. Google AI algorithm masters ancient game of Go. *Nature News* 2016, č.529, s. 445–446.

GRAHAM, Andrea. Coached mobile app platform for the treatment of depression and anxiety among primary care patients: a randomized clinical trial. *JAMA Psychiatry*, 2020, s. 906–914.

GRIFFITH, Caroline. A User Experiences of Social Support From Companion Chatbots in Everyday Contexts: Thematic Analysis, *J Med Internet Res* 2020.

FLORIDI, Luciano. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds & Machines* 2018, roč. 28, s. 689–707.

FRONEMANN, Nora. Should my robot know what's best for me? Human–robot interaction between user experience and ethical design. *AI & SOCIETY*. 2022. s. 517-533.

FULMER, Russel. The Ethics of Psychological Artificial Intelligence: Clinical Considerations. *Conseling and Values* 2021.

FULMER Russell. Artificial intelligence and counseling: Four levels of implementation *Theory & Psychology* 2019, roč. 29, č. 6, s. 807–819.

HOFMANN, Stefan. The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research*, 2012, s. 427-440.

INKSTER, Becky. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR mHealth and uHealth* 2018, roč. 6, č. 11.

KACZKURKIN, Antonia. Cognitive-behavioral therapy for anxiety disorders: an update on the empirical evidence, *Dialogues in Clinical Neuroscience*, 2015, s. 337-346.

KAUER, Silvia. Self-monitoring Using Mobile Phones in the Early Stages of Adolescent Depression: Randomized Controlled Trial. *J Med Internet Res* 2012.

KUŘE, Josef. Prospektivní bioetika, Studie ke konceptu etiky biomedicínských technologií. Praha: Filosofia 2020.

LINDSAY, Geoff. *Etika pro evropské psychology*. Praha: Triton, 2010.

LUXTON, David. Recommendations for the ethical use and design of artificial intelligent care providers, *Artificial Intelligence in Medicine* 2014, s. 1-10,

LUXTON, David D. Ethical implications of conversational agents in global public health. *Bulletin of the World Health Organization* 2020, s. 285 - 287.

MC CARTHY, John. A proposal for the Dartmouth summer research project on artificial intelligence. *AI Magazine* 2006, roč. 4, č 27, s. 12.

NAGATA, Jason. Prevalence and Sociodemographic Correlates of Unmet Need for Mental Health Counseling Among Adults During the COVID-19 Pandemic. *Psychiatric Services* 2022, roč. 73, č.2, s. 206.

NEARY, Martha. State of the Field of Mental Health Apps, *Cognitive and Behavioral Practice*, roč. 25, č. 4, 2018, s. 531-537.

OCISKOVÁ Marie., PRAŠKO, Ján. *Generalizovaná úzkostná porucha*. Praha: Grada, 2017. s. 25.

PATEL, Vimla. The coming of age of artificial intelligence in medicine. *Artificial Intelligence in Medicine* 2009. č. 46, s. 5–17.

POLÁKOVÁ, Jolana. *Smysl dialogu*. Praha: Vyšehrad 2008. s. 57.

PRAŠKO, Ján. *Úzkostné poruchy*. Praha: Portál, 2005.

PROCHASKA, James - NORCROSS, John. *Psychoterapeutické systémy*. Praha: Grada 1996, s. 19.

RÈGUE-GUYON, Mathilde. Neuroepigenetics of Neurotrophin Signaling: Neurobiology of Anxiety and Affective Disorders, *Progress in Molecular Biology and Translational Science*, 2018, s. 159-193.

ROGERS Carl. *Teorie terapie a osobnosti*. Praha: Portál. 2020, s.185.

SEIKKULA, Jaakko. Healing elements of therapeutic conversation: Dialogue as an embodiment of love. *Family Process*, 2005, roč. 44, č.4, s 461–475.

SMITH, Randy. Psychotherapy in the #MeToo era: Ethical issues. *Psychotherapy* 2019, roč. 56, č. 7, s 483-490.

ŠLEPECKÝ, Miloš - PRAŠKO, Ján – KOTIANOVÁ, Antónia – VYSKOČILOVÁ, Jana. *Třetí vlna v kognitivně-behaviorální terapii*. Praha: Portál, 2018.

TIMULÁK, Ladislav. *Základy vedení psychoterapeutického rozhovoru*. Praha: Portál, 2005.

VRÁNA, Karel. *Dialogický personalismus*. Praha: Zvon 1996. s. 22.

Internetové zdroje

An Encyclopædia Britannica Company *MerriamWebster dictionary*. [2022-05-31]. <<https://www.britannica.com/topic/Merriam-Webster-dictionary> > .

BARNDEN, John. Metaphor and Context: A Perspective from AI. (2009) [2022-05-31].
<<http://www.czso.czhttps://www.researchgate.net/publication/304736287> >.

Department for Business, Energy and Industrial Strategy, Industrial Strategy: Building a Britain fit for the future (November, 2017) [2022-05-31].
<https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf> .

Ethical Principles of Psychologists and Code of Conduct. Washington DC: American Psychological Association (2020) [2022-05-31].
<<https://www.apa.org/about/policy/ethics>> .

Ethics guidelines for trustworthy AI. (8.4.2019) [2022-05-31]. <<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>> .

IBM cloud. *AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference?* (27.5.2020) [2022-05-31]. <<https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>>.

International Statistical Classification of Diseases and Related Health Problems, Geneva: WHO, (1.1.2022) [2022-05-31].
<<https://www.who.int/standards/classifications/classification-of-diseases>>.

MAULDIN, Michael. ChatterBots, TinyMuds, and the Turing test: Entering the Loebner prize competition. (1994) [2022-05-31].
<<https://www.aaai.org/Papers/AAAI/1994/AAAI94-003.pdf> >.

National Institute of Mental Health. Depression in College Students. (2017) [2022-5-31]. <<https://www.nimh.nih.gov/health/publications/depression-and-college-students/index.shtml> > .

OPPY, Graham - DOWE David , "The Turing Test", The Stanford Encyclopedia of Philosophy (2021) [2022-05-31] . <<https://plato.stanford.edu/archives/win2021/entries/turing-test/>>.

Woebot CEO Michael Evers on AI in mental health, and how to get a chatbot to bond with a human. (30.7.2021) [2022-07-25]. < <https://vator.tv/news/2021-07-30-woebot-ceo-michael-evers-on-ai-in-mental-health-and-how-to-get-a-chatbot-to-bond-with-a-human>>.

Policy and investment recommendations for trustworthy Artificial Intelligence. (26.6.2019) [2022-05-31]. < <https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence> >.

Recommendation of the Council on Artificial Intelligence. (2019) [2022-5-31]. <<https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>>.

PŘÍHODA, Petr. *Etika : Čtvero základních principů lékařské etiky.*(23.11.2011) [2022-5-31]. <<http://www.lf2.cuni.cz/ustav-lekarske-etiky-a-humanitnich-zakladu-mediciny-2-lf-uk/etika>>.

RASKIN, Hannah. Towards Empathetic Open-domain Conversation Models: A New Benchmark and Dataset. Association for Computational Linguistics. 2019. s. 5370–5381. Responsible AI practices. [2022-05-31]. <<https://ai.google/responsibilities/responsible-ai-practices/> >.

WEIZENBAUM, John. Computer power and human reason: from judgment to calculation San Francisco : W.H. Freeman. (1976) [2022-05-31].

<https://www.researchgate.net/publication/286058724_Computer_Power_and_Human_Reason_From_Judgment_to_Calculation>.

Woebot: A Professional Review (30.8.2019) [2022-5-31].

<<https://onemindpsyberguide.org/expert-review/woebot-an-expert-review/>>.

Woebot CEO Michael Evers on AI in mental health, and how to get a chatbot to bond with a human. (30.7.2021) [2022-07-25]. < <https://vator.tv/news/2021-07-30-woebot-ceo-michael-evers-on-ai-in-mental-health-and-how-to-get-a-chatbot-to-bond-with-a-human>>.