UNIVERZITA KARLOVA

FAKULTA SOCIÁLNÍCH VĚD

Institut sociologických studií

Katedra sociologie

Diplomová práce

2021 **Jakub Janouš**

UNIVERZITA KARLOVA

FAKULTA SOCIÁLNÍCH VĚD

Institut sociologických studií

Katedra sociologie

Zdroje důvěry v inteligentní virtuální asistenty

Diplomová práce

Autor práce: Jakub Janouš

Studijní program: Sociologie

Vedoucí práce: doc. PhDr. Dino Numerato, Ph.D.

Rok obhajoby: 2021

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 dne 1.1. 2021 Jakub Janouš

Bibliografický záznam

JANOUŠ, Jakub. *Zdroje důvěry v inteligentní virtuální asistenty.* Praha, 2021. 91 s. Diplomová práce (Mgr.). Univerzita Karlova, Fakulta sociálních věd, Institut sociologických studií. Katedra Sociologie. Vedoucí diplomové práce doc. PhDr. Dino Numerato, Ph.D.

Rozsah práce: 205 923 znaků

Anotace

Diplomová práce se zabývá zdroji důvěry v umělou inteligenci, a jak jsou tyto zdroje podmíněné sociálními reprezentacemi. Zkoumá, jaká rizika si uživatel uvědomuje při používání umělé inteligence a jak různé faktory ovlivňují uživatelovo smýšlení. Umělá inteligence je v mé práci reprezentovaná inteligentními virtuálními asistenty (IVA). Na základě polo-strukturovaných výzkumných rozhovorů s jejich uživateli jsem jako hlavní zdroje důvěry určil: neutralitu, víru v budoucnost, naplnění očekávání a blízkost, přičemž první dva zdroje jsou stálé a vychází ze sociální struktury, zatímco poslední dva zdroje důvěry vychází přímo ze zkušeností uživatele a jsou proměnlivé v čase. Dále bylo zjištěno, že sociální reprezentace mají významný vliv na zdroje důvěry. Veškeré sociální reprezentace jsem rozdělil do tří kategorií – strojovost, personifikace, nehmatatelnost, dle kterých se dalo posuzovat uživatelovo vnímání umělé inteligence a na základě toho i jeho uvažování o ní. Díky tomu jsem dokázal, že důvěra uživatelů je sociálním reprezentacemi podmíněna. Důležitou součástí důvěry je uvědomování si rizik. V rámci mé práce jsem určil šest hlavních rizik, které si uživatelé uvědomují: nepředpokládaná chyba softwaru, mechanická chyba, zneužití IVA třetí stranou, právní rizika, strach z odcizení a ovládnutí technologiemi. V neposlední řadě se práce zabývá vlivem sociálních, politických, ekonomických a kulturních faktorů, které ovlivňují tvorbu sociálních reprezentací a tím i uvažování a důvěru v umělou inteligenci.

Annotation

The diploma thesis deals with the sources of trust in artificial intelligence, and how these sources are conditioned by social representations. It examines what risks the user is aware of when using artificial intelligence and how various factors affect the user's thinking. Artificial intelligence is represented in my work by intelligent virtual assistants (IVA). Based on semi-structured research interviews with their users, I have identified as the main sources of trust: neutrality, belief in the future, fulfillment of expectations and closeness. The first two sources are constant and based on social structure, while the last two sources of trust are based directly on user experience and are variable over time. It was also found that social representations have a significant impact on sources of trust. I divided all social representations into three categories - mechanicality, personification, intangibility, according to which the user's perception of artificial intelligence could be assessed and, based on that, his thinking about it. Because of this, I have proved that the trust of users is conditioned by social representations. An important part of trust is risk awareness. In my work, I have identified six main risks that users are aware of: unexpected software error, mechanical error, misuse of IVA by a third party, legal risks, fear of alienate and control by technology. Last but not least, the work deals with the influence of

social, political, economic and cultural factors that affect the creation of social representations and thus thinking and confidence in artificial intelligence.

Klíčová slova

Teorie sociálních reprezentací, důvěra, důvěra v umělou inteligenci, zdroje důvěry, digitální sociologie, automatizace, umělá inteligence, inteligentní virtuální asistent

Keywords

Theory of social representations, trust, trust in artificial intelligence, sources of trust, digital sociology, automation, artificial intelligence, smart virtual assistant

Title/název práce

Sources of trust in intelligent virtual assistants

Poděkování

Rád poděkoval panu doc. PhDr. Dino Numeratovi, Ph.D. za odborné vedení, konzultace a velmi vstřícný, přátelský a obětavý přístup. Jsem velmi vděčný za jeho připomínky, návrhy a věcné podněty, díky kterým mohla vzniknout tato práce. Také děkuji panu Marku Němcovi z WS Czech za poskytnutí tiskové zprávy IBM pro akademické účely. V neposlední řadě děkuji svým rodičům za jejich podporu a motivaci nejen při psaní své diplomové práce, ale v průběhu celého studia.

Obsah

Úvod

„*Bixby, are you artificial intelligence?“*

„I don’t think it’s possible to sum up what I am in a single word.“[1]

Není jednoduché určit, zda inteligentní virtuální asistenti jsou skutečnou umělou inteligencí v současném světě, ale na základě jejich schopností mají k umělé inteligenci minimálně velmi blízko.

Umělá inteligence (UI) je v různých podobách běžnou součástí života mnoha lidí, a to aniž by si to často uvědomovali. Lidé s ní interagují ať už přímo, či nepřímo, když něco hledají na internetu (v prohlížeči), je možné ji nalézt ve zdravotnictví, v průmyslu, ve formě chatbotů v poradenství, v počítačových hrách, či právě v mobilním telefonu ve formě inteligentního virtuálního asistenta. [Sun, Medaglia 2019: 368-383] Umělá inteligence je všude kolem nás. Je to právě umělá inteligence, která je implementována do internetových vyhledávačů jako je Google Chrome, či Firefox a rozhoduje o tom, který článek se nám objeví jako první. Je také součástí sociální sítě Facebook. Umělá inteligence rozhoduje, kdy a jakým výtahem pojedeme, může pomoci v nemocnici při diagnostice, může řídit pracovní činnosti ve firmě, může pomáhat lidem při řešení drobné dopravní nehody, anebo je přepojit na konkrétní specializované pracoviště. [Russell, Norwig 2010: 1–17] Přesto, jaký vliv má umělá inteligence na životy lidí, nevěnuje ji sociologie takovou pozornost. Zejména výzkumů, které se věnují názorům uživatelů a specifikům přístupu k umělé inteligenci, je velmi málo. [Sun, Medaglia 2019: 368-383]

Touto prací jsem se rozhodl přispět k zaplnění bílých míst v sociologickém zkoumání umělé inteligence a pomoci s otázkou, co vlastně je umělá inteligence ve vztahu k člověku a společnosti. Přesněji se budu zabývat tím, proč lidé umělé inteligenci důvěřují – jaké k tomu mají důvody, z čeho důvody pramení a jaká si uvědomují rizika. Umělou inteligenci v mé práci budou nahrazovat právě virtuální asistenti, kteří jsou dnes již běžnou součástí většiny chytrých telefonů. Jejich dovednosti jsou navíc postaveny na schopnosti strojového učení, který je nedílnou součástí umělé inteligence.

V této diplomové práci se tedy soustředím na to, jak lidé o umělé inteligenci uvažují, k čemu ji připodobňují a jak se formuje uživatelova důvěra v umělou inteligenci. Odpovídám konkrétně na čtyři výzkumné otázky, které si pokládám: Jaké jsou hlavní zdroje důvěry v umělou inteligenci

[1]Odpověď inteligentní virtuálního asistenta Bixby od společnosti Samsung na otázku, zda je umělá inteligence. [cit. 27. 2. 2020]

v každodenním životě? Jak je důvěra podmíněná sociálními reprezentacemi o umělé inteligenci? Jaká rizika si jedinec uvědomuje ve vztahu k umělé inteligenci? A nakonec: Jakým způsobem je důvěra v umělou inteligenci ovlivněna sociálními, politickými, ekonomickými a kulturními faktory? K zodpovězení otázek využiji teorii sociálních reprezentací od Serge Moscoviciho. Na základě této teorie, která může být chápana také jako výzkumná metoda, jsem uskutečnil 10 polo-strukturovaných výzkumných rozhovorů s uživateli inteligentních virtuálních asistentů. Jejich odpovědi jsem analyzoval pomocí deduktivní klasifikace výroků a pomocí indukční analytické strategie.

V práci nejprve seznámím čtenáře s teoretickým vhledem do problematiky zkoumaných konceptů. Na začátku představím různé koncepty důvěry v sociálních vědách a diskutuji jejich relevanci pro zkoumání vztahu k umělé inteligenci. Následně představím teorii sociálních reprezentací a ukážu možnosti jejího využití. Teoretickou část uzavřu reflexí technologického vývoje s důrazem na nebezpečí a paradoxy pro uživatele, které z toho plynou. Následuje metodologická část, v níž čtenáře seznámím s designem výzkumu. V analytické části uvedu tři kategorie vnímání virtuálních asistentů, jak se v ně formuje důvěra a jaké jsou její hranice a jaká rizika si uživatelé uvědomují. Své výsledky následně shrnu v závěru a představím další možná témata, jimiž je možno se ve vztahu k virtuálním asistentům, umělé inteligenci, či sociálním reprezentacím dále zabývat.

Teoretická část

Důvěra

Koncept důvěry v pojetí klasických sociologů

Klíčovým konceptem této práce je pojem důvěry a ve vztahu k automatizaci a umělé inteligenci. V první řadě je tedy potřeba si tento pojem přiblížit a pochopit jeho úlohu a podmínky ve společenské interakci. O důležitosti důvěry mluvil například Émile Durkheim, který považoval důvěru jako neodmyslitelnou součást organické solidarity. V důsledku dělby práce ve společnosti docházelo k tomu, že si lidé museli mezi sebou více důvěřovat, aby mohl společenský systém dál fungovat a rozrůstat se. [Sedláček 1979: 74–76] Jako samostatným pojmem se důvěrou poprvé zabýval Georg Simmel, který důvěru popisuje jako důležitou součást společenské interakce. Jedná se o základní kámen společenské komunikace, respektive o základní kámen sociálních vztahů, které komunikaci podporují. [Simmel 1997: 178] Nicméně Simmel o důvěře uvažoval pouze jako o vztahu mezi lidmi.

Luhmann upozorňuje, že je potřeba rozlišovat mezi pojmy důvěra a důvěřivost (v originále: trust and confidence). Ten podobně jako Simmel, zkoumá důvěru v kontextu mezilidské komunikace.[2] Luhmann popisuje *důvěru* (*trust*) jako očekávání určitého chování, přičemž si jedinec uvědomuje možnost rizika. Rizikem může být myšleno nesplnění očekávaného chování, nebo důsledky plynoucího z nesplnění očekávaného chování, přičemž jsou rizika vždy orientována na konkrétní prostředí a čas. Naopak *důvěřivost* (*confidence*) je slepá víra v něco. Jedinec v takovém případě neuvažuje o možných rizicích a nepřipouští si žádné možnosti nebezpečí plynoucí ze vzájemné interakce. Důvěřivost může být vědomá i nevědomá. [Luhmann 1979: 18–22] Erik Eriksson toto rozdělení ještě dále rozvíjí. Důvěřivost vnímá jako víru, která stojí na počátku jakéhokoliv vztahu. Jelikož jedinec nemá žádné předchozí zkušenosti, je nucen slepě věřit. Eriksson to dokládá příkladem malých dětí, jejichž důvěra v rodiče je spíše vírou v ně.[3] Důvěru (trust) vnímá jako očekávání, či spoléhání se na něco, vyplývající z předchozí vzájemné zkušenosti. [Eriksson 2002: 227]

Naproti tomu se však staví Giddens. Nesouhlasí totiž s Luhmannovým tvrzením: „…když se zdržíš jednání, nevystavuješ se riziku – jinými slovy, že kdo nic nedělá, nic nezkazí.“ [Giddens 2003: 36] To považuje za nesprávný výklad důvěry, neboť považuje důvěru za kontinuální proces, nerozlučně

[2] Luhmann mimo jiné píše o důležitosti konceptu důvěry v kontextu veškerých sociálních interakcí ve vztahu k moderním technologiím. [Luhmann 1979: 18–22]

[3] Eriksson však přiznává, že nejedná čistou víru, nebo důvěřivost. I zde je určité uvědomování si rizika. Nicméně vnímané riziko je natolik nepodstatné pro interakci, že pojem důvěřivost se na tento příklad hodí. Eriksson v tomto ohledu také píše o *naivní důvěře*.

spjatý se sociálními strukturami, který podporuje jejich fungování, přičemž je oproštěn od času i prostoru. Důvěřivost vnímá jako tzv. „*slabé induktivní vědění*“. Tím upozorňuje na fakt, že důvěra je v určitém smyslu vždy slepou důvěrou. Důvěřivost vyjadřuje víru v poctivost nebo lásku druhého, či v přesnost abstraktních principů (technických znalostí). [Giddens 2003: 33–39] Giddens si pokládá otázku: „Proč většina lidí zpravidla důvěřuje praktikám a sociálním mechanismům, o nichž má jen nepatrné, nebo žádné technické vědomosti?“ [Giddens 2003: 82] A sám si na ní také odpovídá: Důvěra je tam, kde je nevědomost. Věcem, které jedinec dokonale zná, důvěřovat nemusí, protože dokáže odhadnout každou následující akci. Ale které věci má možnost jedinec tak dokonale znát, aby takto mohl určit každý následující krok? Tam, kde končí znalosti tak přichází důvěra. Tato důvěra vychází z vědy a technických znalostí. Giddens uvádí příklady, kdy jsou jednotlivé vědní disciplíny považovány laiky za studnici nepochybných pravd. Jelikož laik o dané vědě nic neví, je odkázán na vědomosti lidí, kteří se tím zabývají, a tím pádem jim důvěřuje (důvěra v *expertní systémy*). [Giddens 2003: 75–101] V post-faktické společnosti[4] navíc dochází k oslabení role faktických vědeckých vysvětlení. Současné vědecko-populární proudy dávají přednost silným a přesvědčivým příběhům před holou pravdou. Následkem je změna odborného a vědeckého informování, které má vliv na rozhodování společností. Berling a Bueger ve svém textu dokonce hovoří o devalvaci vědeckých a odborných znalostí. [Berling a Bueger 2017: 1–10] To však ještě posiluje systémovou stereotypizaci vědy a technických poznatků, kdy jsou vědci a technici bráni laiky jako odborníci ve svém oboru, kteří „to“[5] musí přece vědět. Lidé jsou tak v mnoha případech odkázáni na důvěru v expertní systémy. Důvěra je totiž vynucována samotným systémem. Pokud chce jedinec žít v dané společnosti, musí přijmout fungování jeho institucí. Jelikož však o něm jedinec nemusí mít stoprocentní znalosti, je jedinec nucen institucím důvěřovat – koncept důvěry v *abstraktní systémy*. [Giddens 2003: 75–101]

Teoreticky tak lze říci, že důvěra v umělou inteligenci vychází z dvojího předpokladu. Prvním předpokladem je, že vývojáři umělé inteligence jsou odborníci, kteří naprogramovali funkční technologii, která má lidem usnadnit život. Laici, pokud nejsou sami programátoři umělé inteligence, si tuto skutečnost ověřit nemohou, a tak musí často důvěřovat/věřit ve správné naprogramování UI – (důvěra v *expertní systémy*). Druhým předpokladem je, že umělá inteligence má schopnost

[4] Post-faktická společnost (post-factual někdy také post-truth) se staly pojmy používanými k popisu současných posunujících se vztahů mezi vědeckými znalostmi a politikou. Od roku 2016 je možno dohledat pojem „Post-truth“ také Oxfordském anglickém slovníku. [Berling a Bueger 2017: 2]

[5] Jako příklad můžu uvést situaci, kdy se v novinách píše, že vědci objevili lék na rakovinu. V tuto chvíli laik může nabýt dojmu, že lék je vytvořen a připraven k distribuci – důvěřuje vědcům. Lékaři jsou však skeptičtí, neboť mají větší znalosti o daném oboru a vědí, že lék se musí ještě několik let testovat, než bude vpuštěn do oběhu.

strojového učení, a tak se daná technologie sama zdokonaluje. Za „odborníka“ lze v tomto případě považovat samotnou UI. Uživatelé mohou věřit v její schopnosti učení a schopnosti účelného rozhodování. To je však pouhá teorie, na kterou částečně reaguje a odpovídá tato práce.
Jelikož koncept důvěry tvoří nezanedbatelnou část sociologického bádání, lze důvěru zkoumat z různých perspektiv. Řada autorů se příliš neodlišuje od předchozích pojetí a vnímají důvěru často ve vztahu k očekávatelnému chování někoho, nebo něčeho. Například Rotter uvádí, že důvěra je založená na předem udělenému slibu, informaci, nebo jiné aktivní formě komunikace. [Rotter 1967: 651–652] Johns vnímá důvěru: „jako ochotu jít do vztahu, při kterém se vytváří nebo zvyšuje zranitelnost se spoléháním se na někoho nebo na něco, co se bude prezentovat podle očekávání.“ [Johns 1996: 81] Je také možno zkoumat jaké faktory mají na důvěru vliv a co se stane při jejím porušení. [Rousseau, Sitkin, Burt, Camerer 1998: 393–402] Toto je však pouze jeden úhel pohledu, který nemusí být pro zkoumání důvěry v umělou inteligenci dostačující.

Dalším možným úhlem pohledu je zkoumání důvěry ve vztahu k času. Myšlenka je založena na předpokladu, že důvěra není konstantní, ale proměňuje se v čase. Z důvodu komplexnosti a širokého záběru jsem vybral článek od Roye Lewickiho a Barbary Bunker, kteří důvěru rozdělili do tří fází. V 1. fázi je důvěra založená na nedostatku informací a člověk je odkázaný na důvěru v expertní vědění. Ve 2. fázi je důvěra založená na znalostech (zkušenosti), na jejímž základě dochází k prohloubení důvěry. Ve 3. fázi je důvěra založená na sociální identifikaci – člověk důvěřuje, neboť se s danou entitou (člověkem) identifikuje. [Lewicki, Bunker 1995: 133 – 173] Autoři se od předchozích pojetí odlišují temporálním rozdělením intenzity důvěry. Nicméně se nabízí otázka, jak se slučuje s důvěrou v umělou inteligenci. Je například možné, aby člověk navázal tak silnou důvěru k uměle vytvořené věci, že by se s ní identifikoval (viz 3. fáze důvěry)? Či je možné, aby UI mělo takové schopnosti, aby se s ní lidé dokázali identifikovat? Na tyto otázky patrně v této práci odpovědět nedokáži. Dokáži ale nastínit myšlení, jak nad důvěrou v umělou inteligenci uvažovat.

Existují i další dimenze důvěry, které může badatel dále zkoumat. Předchozí autoři se zabývají především orientací na zkušenosti, či temporalitou. Frederiksen popisuje hned tři další různé dimenze, které je možné v souvislosti s důvěrou zkoumat. Jsou to: *dimenze vztahů*, *dimenze objektů* a *dimenze situací*.

I. *Dimenze vztahů* je založená na blízkosti a typu vztahů. Dělí se na blízkost slabá (kolega z práce) střední (někteří členové rodiny, kamarádi) a vysokou (nejbližší rodina a blízcí přátelé. Dimenze vychází ze vzájemné reciprocity dostupnosti pomoci.
II. *Dimenze objektů* znamená, že badatel zjišťuje, v čem člověk někomu důvěřuje (záleží na objektu). Frederiksen uvádí, že přestože dle první dimenze si jsou otec s dcerou blízcí (vysoká blízkost), otec nemusí dceři důvěřovat a nepůjčí ji peníze. Na druhou stranu je

ale například půjčí někomu jinému, kdo není tak blízký. Frederiksen zde poukazuje, že každému věříme jinak v jiných věcech a od toho se také odvíjí míra důvěry.

III. *Dimenze situací* závisí na kontextu. Zde je důležité, jaký vliv má důvěra na vzájemný vztah nejen v současné chvíli, ale také jaké je očekávání do budoucna. Frederiksen upozorňuje na silný vliv této dimenze, která ovlivňuje vzájemnou důvěru. [Frederiksen 2012: 733 – 750].

Jak je vidět, důvěru lze pojímat z různých úhlů pohledů a z různých dimenzí. Většina z nich vychází ze zkušenosti – zda dojde, či nedojde k naplnění nějakého očekávání a tím k nastolení nějaké míry důvěry. Na tomto základě se může důvěra proměňovat v čase, kdy se mění její intenzita. Podoba důvěry však není závislá pouze její intenzitě, či na naplnění očekávání, ale také na vzájemném vztahu mezi subjekty, co je předmětem vztahu důvěry a na celkovém kontextu. Otázkou je, jak jsou tato pojetí důvěry slučitelné s důvěrou lidí v umělou inteligenci, která není živoucí entitou, ale je uměle vytvořena člověkem? Důvěra je obvykle zkoumána ve vztahu dvou a více osob, maximálně k nějaké instituci, která má živoucí charakter (škola, úřad). Jaká je tedy podoba důvěry technologii, která dokáže na člověka aktivně reagovat a na základě okolních podnětů utvářet vlastní rozhodnutí?

Důvěra v automatizaci a umělou inteligenci

Důvěra jedince v automatizaci je jiná než mezilidská důvěra. Lee a See to vysvětlují tím, že automatizace jako taková postrádá vlastní úmysl, a tak má člověk tendenci jí věřit. [Lee, See 2004: 65] Na druhé straně autoři zde zapomínají na úmysl „stvořitele“ dané automatizace, tedy úmysl toho, kdo ji vytvořil a naprogramoval. Každý algoritmus tak nese stopu „lidského úmyslu“. Přesto mají někteří lidí tendenci přistupovat k technologiím neutrálně. Příkladem může být zpracovávání a analýza tzv. Big Data, která poskytují iluzi objektivních a exaktních informacích, která se můžou lišit od následného zpracování, mohou být vytržena z kontextu, většinou jsou různě upravována a ořezávána, a především jsou mylně chápána jako ekvivalent sociálních sítí, či sociálních interakcí v reálném světě. [Boyd, Crawford 2012: 662–679] Lidé si mohou tvořit o technologiích a automatizaci různé iluze, které nemusí korespondovat s realitou. To právě závisí na různých sociálních reprezentacích, které si jedinec vytváří.

O dalším důkazu iluzorních pohledů lidí na technologie píší Madhavan a Wiegmann. Ti rozdělili hlavní odlišnosti důvěry ve vztahu jedinec a jedinec a důvěry jedinec a stroj do tří kategorií, podle kterých reagují lidé jinak na jiné lidi a jinak na stroje.

1. *Spolehlivost stroje* – vše plyne z přesvědčení, že stroje jsou v každé specializované činnosti lepší než lidé. Stroj je vnímán lidmi jako odborný specialista s odpovídajícími schopnostmi přesně určený ke konkrétní činnosti.

2. *Důvěryhodnost* – lidé mají tendenci strojům věřit více než jiným lidem. Jedinec si tak méně uvědomuje potenciální rizika chyby stroje než rizika chyby, kterou může udělat jiný člověk. Varování před chybou stroje jsou uživateli mnohem více přehlížena než varování před chybou jiného jedince. Tato diverzita je zapříčiněna předpokladem[6] vyšší odbornosti technologie. [Madhavan, Wiegmann 2007: 281–288] Zda tomu tak skutečně je mohou ohladit právě sociální reprezentace.
3. *Zdroj diagnostických operací* – lidé jsou mnohem vnímavější k chybám strojů. Jedinec neočekává, že u stroje nastane chyba. U lidí se do jisté míry chybovost předpokládá, a chyby si tak jedinci nepamatují. U stroje se ale tolik chyb neočekává a tím pádem má uživatel větší tendenci si chyby technologie pamatovat, neboť je nepředpokládá. [Madhavan, Wiegmann 2007: 281–288]

Tato teorie je v celku jednostranná a poukazuje na idealistické pojetí uživatelů automatizace. Pomocí teorie sociálních reprezentací zjistím, jak lidé skutečně uvažují o chybách a rizicích spojených se vzájemnou interakcí. Přesto sloučením konceptů Boyda, Crawforda a Madhavana s Wiegmannem dostaneme definici důvěry v automatizaci, která je chápána jako sociálně psychologický koncept, kdy agent pomůže dosáhnout cíle jednotlivce, v situaci, která je charakterizována nejistotou a zranitelností. [Lee, See 2004: 55] „Agentem se v oboru umělé inteligence označuje každá entita, která dokáže vnímat své prostředí (prostřednictvím senzorů, či čipů), působit na něj a reagovat. “ [Shapiro 1992: 641] Tato definice nicméně neobsahuje dimenze a koncepty z předchozí části, které jsou však, dle mého názoru, stejně relevantní pro umělou inteligenci, stejně tak jako pro mezilidské vztahy.

O existenci specifické důvěry v kybernetickém prostředí píší také autoři Han a Sang-Lin, kteří se zabývají pojmem *e-trust*[7]. Jejich pozornost se ubírala na internetové obchody a na faktory, které ovlivňují důvěru lidí zde nakupovat.[8] Výsledkem bylo zjištění, že e-trust v nakupování přes internet ovlivňuje především kvalita webových stránek a opakovaná interakce [Han, Sang-Lin 2007: 101 – 122]. Jejich výzkum byl však omezen na důvěru ve vztahu ke splnění očekávání, či dobré zkušenosti, přičemž nebrali v úvahu i další dimenze, kterých může důvěra nabývat. Podobný přístup zvolili i autoři Liu, Yang a Xu, kteří zkoumali faktory přijetí nové technologie. Důvěra je, dle jejich výzkumu,

[6] Úsudek uživatele není založen na důkladném prozkoumání technologie. Je založen na povrchových charakteristikách zdroje informací o dané technologii – tedy odbornosti a důvěryhodnosti. [Madhavan, Wiegmann 2007: 285]

[7] Neexistuje všeobecně uznávaná definici.

[8] Paralela s mým výzkumem zaměřený na virtuální asistenty a na faktory, které ovlivňují důvěru je používat.

ovlivňována dvě druhy vlivů: přímými a nepřímými účinky. Nepřímé účinky jsou častější o kognitivní cestu sociální důvěry, kdy všeobecné přijetí technologie dá důvěru jedinci danou technologii použít. Mnohem důležitější a také účinnější je přímý vliv, tzv. afektivní cesta. Tedy situace, kdy sám jedinec danou technologii vyzkouší a sám k ní získá důvěru [Liu, Yang, Xu: 2018] Tento poznatek je pro moji práci velmi důležitý, neboť potvrzuje určitou hierarchičnost jednotlivých faktorů a rozdílnou intenzitu jednotlivých vlivů.

Hengstler, Enkel a Duelli zvolily pro své bádání jiný směr. Rozhodly se zkoumat, jak je důvěra v umělou inteligenci podporována distributorem technologie, či komunikací firmy. Jejich hlavní myšlenkou je, že v kontextu inovací má velký význam zvažování vnímaného rizika, protože je ústředním proudem pro přisvojení technologie uživatelem. [Hengstler, Enkel a Duelli: 2016: 106] To koresponduje s tezí Ulricha Becka, který považuje vnímaná rizika za zprostředkovaná. Vychází z myšlenky, že se potenciální rizika moderní společnosti vymykají možnostem lidského vnímání. Jedinec si těžko představí, co přesně znamená zvýšený podíl oxidu uhličitého v atmosféře a co to pro něj znamená. [Beck: 2018: 35] Autorky docházejí k závěru, že pro důvěru uživatele v technologii je zásadní seznámení uživatele s danou technologií a silná podpora důvěry ze strany výrobce – aktivní komunikace směrem k zákazníkovy a snaha řešit jeho problémy. [Hengstler, Enkel a Duelli: 2016: 106–114] Toto zjištění však může být zavádějící, neboť jak ukáži v následujícím odstavci, lidé mohou důvěřovat technologii i bez předchozí interakce, či komunikace s distributorem.

Asi nejkomplexnější výzkum důvěry v automatizaci provedli Hoff a Bashir. Ti udělali rešerši 101 empirických prací skládajících se ze 127 samostatných studií. Na konci šetření určily, že důvěra v automatizaci má tři vrstvy. První vrstvou je *dispoziční důvěra*. Dispoziční důvěra představuje trvalou tendenci jednotlivce důvěřovat automatizaci, vyplývající z biologických a environmentálních vlivů. Druhou vrstvou je *situační důvěry*, kdy důvěra vychází z konkrétní situace a jsou vnímána potenciální rizika a výhody. Autoři zde mluví o situacích, kdy člověk je k odkázaný na důvěru v automatizaci. Tak upozorňují, že často dochází k tomu, že u vysoce rizikových činností, se snižuje frekvence využívání automatizace. Třetí vrstvou je *naučená důvěra.* Jedná se o důvěru, jejíž intenzita kolísá v průběhu používání v závislosti na zkušenosti. [Hoff, Bashir: 2014] Dle mého názoru je však druhá a třetí vrstva důvěry silně provázaná, respektive jejich hranice nejasná. Nemůže být situační důvěra podmíněná předchozími zkušenostmi? Dispoziční důvěra je závislá na kultuře, věku, pohlaví a osobnosti, což jsou, může být pravda, na druhou stranu je tato charakteristika poměrně obecná.

V roce 2018 byl v České republice proveden výzkum od MNS Market Research pro IBM zabývající se důvěrou společnosti v umělou inteligenci. Z tohoto výzkumu vyplynulo, že vědomou osobní zkušenost s umělou inteligencí implementovanou do každodenního života má 19 % respondentů.

T tom samém průzkumu odpovědělo 56 % dotazovaných, že mají důvěru v umělou inteligenci. To znamená, že minimálně 37 % dotazovaných důvěřuje umělé inteligenci, i když s ní nemá vědomě žádné zkušenosti. [IBM – Artificial Intelligence: 2018][9] Tento průzkum probíhal také Polsku, Maďarsku a Rusku, a to s velmi podobným výsledkem. Toto zjištění je však v rozporu s tvrzením autorek, že pro důvěru v technologie je zásadní seznámení uživatele s danou technologií. To do jisté míry uvádí i Beck, který upozorňuje na rozdílnost vědecké a sociální reality. Přesto že, jedinci potřebují k získání důvěry odborný názor, často se jim očekáváné odpovědi nedostane, nebo se jim nedostane v té podobě, v jaké by ji potřebovali. Laici nemusí vždy rozumět odborným termínům a hlavně nemají potřebné zkušenosti s danou technologií. Právě tak vznikají trhliny mezi sociální a vědeckou realitou. [Beck: 2018: 39] Uživatelé tak mnohdy spoléhají na odborné znalosti vědců a důvěřují tak v expertní systémy. [Giddens 2003: 75–101] Avšak i toto pojetí zcela nepodporuje zjištění autorek o vztahu mezi mírou osobní zkušenosti uživatele s danou technologií a mírou důvěry v ní.

Tabulka 1: ***Přehled autorů a jejich hlavní teze.***

Jméno autora	Hlavní teze	Pojetí důvěry	Původ důvěry
Georg Simmel	Základ společenské komunikace	Podporuje sociální vztahy a komunikaci mezi lidmi.	Podmínka komunikace
Niklas Luhmann	Důvěra / Důvěřivost	Uvědomování si možného rizika ze vzájemné interakce.	Vychází z vědomí možného rizika
Anthony Giddens	Slepé induktivní vědění	Důvěra je tam, kde je nevědomost – kontinuální důvěra v abstraktní systémy.	Vychází ze systémů
Julian Rotter	Očekávané chování	Důvěra vychází z předchozího zkušenosti s druhou stranou.	Vychází od „*druhého*“
Hoff a Bashir	Dispoziční důvěra Situační důvěra Naučená důvěra	Důvěra v automatizaci se projevuje ve třech vrstvách.	Predispozice důvěřovat Z konkrétní situace Ze zkušeností
Lewicki, Bunker	Tři fáze důvěry	Důvěra se v čase proměňuje.	Vychází z informací
Morten Frederiksen	Tři dimenze důvěry	Důvěra se různí podle dimenzí – vztahy, objekty, situace.	Původ důvěry se liší napříč dimenzemi.
Madhavan a Wiegmann	Důvěra mezi lidmi a důvěra ke strojům je rozdílná.	Lidská důvěra ve stroje je jiná než důvěra v jiného člověka.	Vychází z kombinace Slepého induktivního vědění a důvěřivosti.
Han a Sang-Lin	E-trust	Důležitost naplnění očekávání a vzbuzení pocitu kvality.	Vychází z pocitu odbornosti a kompetentnosti.
Liu, Yang, Xu	Dva druhy vlivů na přijetí technologie	Přímý vliv je vždy účinnější než zprostředkovaná zkušenost.	Vychází ze zkušenosti.

[9] Průzkum byl vytvářen pro IBM a většina obsahu veřejná. Informace čerpám z tiskové zprávy.

Teorie sociálních reprezentací

Teorie sociálních reprezentací dle Serge Moscovice

Sociální reprezentace je poměrně komplexní pojem. Tato teorie totiž leží na pomezí psychologie a sociologie, neboť sám Serge Moscovici byl sociální psycholog. Svoji teorii tedy koncipoval jako nástroj zabývající se tím, jak lidé přebírají různé informace a jak si je ukládají. Zabýval se především kognitivní složkou, která jedinci restrukturalizuje realitu. [Moscovici 1988: 211–250]

Pro účely této práce jsem se rozhodl využít definici sociálních reprezentací z pojednání *Notes Towards a Description of Social Representations* [Moscovici 1988], která je původní Moscoviciho definice, a která mi pomůže rozkrýt, jakým způsobem jedinec uvažuje o umělé inteligenci a také se zabývá, jakým způsobem poté jedinci s těmito myšlenkami pracují. Definice teorie sociální reprezentace zní: „Sociální reprezentace se zabývá obsahem každodenního myšlení a zdroji myšlenek, které nám dávají soudržnost v našem náboženském vyznání, politickým myšlenkám a souvislostem, které spontánně vytváříme, tak jako když dýcháme. Umožňují nám klasifikovat osoby a předměty, porovnávat a vysvětlovat chování a objektivizovat je, jako součást našeho sociálního prostředí. Nacházejí se nejenom v mysli mužů a žen, ale nalezneme je tak ‚ve světě' a jako takové mohou být zkoumány."[10] [Moscovici 1988: 214]

Sociální reprezentace však nelze považovat za jasné logické a koherentní myšlenkové vzorce. [Höijer 2011: 5] Jejich vytváření a osvojení je složitý proces, do kterého vstupuje mnoho faktorů. Podílejí se na společném vnímání světa, které si společenské skupiny vytváří samy pro sebe. [Moscovici 1988: 217–225] Tato teorie vlastně specifikuje, jak se formuje kolektivní vědomí ve společnosti a jak je následně toto vědomí transformováno prostřednictvím socio-kognitivních procesů a mechanismů do celkového obrazu a vnímání světa. [Höijer 2011: 6] Jde o proces, při kterém si člověk vytváří náhled na realitu na základě přijímaných informací a nějak ji následně interpretuje. Důležité však je, že všechny sociální reprezentace mají za cíl přiblížit jedinci, sociální skupině, či nějaké společnosti neznámý pojem (neznámý objekt). [Moscovici 1988: 228–238] Právě ona neznámost hraje zásadní roli. Neznámost a potřeba vysvětlení neznámého je hlavní důvod vzniku teorie sociálních reprezentací. Popisuje a vysvětluje, jakým způsobem lidé přebírají informace o neznámém prvku v jejich sociální realitě a jakým způsobem s ním dále nakládají. V kontextu této práce je umělá inteligence považována za neznámý pojem, o jejíž vlastnostech si lidé mohou vytvářet mylné

[10] Originální text: Social representations (…) concern the contents of everyday thinking and the stock of ideas that give coherence to our religious beliefs, political ideas and the connections we create as spontaneously as we breathe. They make it possible for us to classify persons and objects, to compare and explain behaviours and to objectify them as part of our social setting. While representations are often to be located in the minds of men and women, they can just as often be found "in the world", and as such examined separately.

představy, či jí přisuzovat určité vlastnosti.

Proces přebírání a vysvětlení neznámých entit rozdělil Serge Moscovici na dvě hlavní fáze. V první fázi dochází k *ukotvení*. Jedná se o proces, kdy jedinec pojmenovává určitou neznámou entitu. Tím se daná skutečnost stala součástí sociální reality jedince. Höijer ukotvení připodobňuje kulturní asimilaci. Nové sociální reprezentace jsou přiřazeny k známým reprezentacím (známým skutečnostem, které jedinec zná už z dřívější doby). Reprezentace jsou tak soustavně ukotvovány a transformovány novými informacemi a prožitky, což ovlivňuje jak nové sociální reprezentace, tak i ty staré. [Höijer 2011: 7–12]

Druhá fáze procesu přebírání a vysvětlení neznámých entit se nazývá *objektivizace*. Pro jedince je tato druhá fáze složitější, neboť zatímco v první fázi *ukotvení* se nové vjemy (potažmo reprezentace) ukládaly podvědomě, ve druhé fázi *objektivizace* musí již jedinec vyvinout určité úsilí. Ve druhé fázi dochází v mysli jedince k integraci nových informací do systémů známých kategorií. Tím si jedinci vlastně interpretují realita. Jedinec si tímto procesem často vysvětluje situace, nebo věci, které nezná, a dobírá se k nim prostřednictvím jemu známých prvků. [Moscovici 1988: 228–238] Jinými slovy si jedinec k abstraktnímu pojmu připojí konkrétní vlastnost. To způsobí, že abstraktní pojem se transformuje do objektivního pojmu ukotveném v realitě, se kterým již jedinec dokáže dále pracovat. [Höijer 2011: 12–13]

Další pojetí teorie sociálních reprezentací

Jedním z autorů, kteří rozdělují sociální reprezentace podle jejich struktury nebo organizace jsou Wanger a Valencia. Ti rozlišují mezi dobře strukturovanými reprezentacemi a těmi méně dobře strukturovanými kognitivními procesy. Podobně jako Moscovici rozdělují[11] reprezentace do dvou kategorií – jádro a periferie.[12] Všímají si, že sociální reprezentace mají určitou formu a strukturu, podle které se dají reprezentace kategorizovat. „…*centrální jádro* je konstitutivní součástí sociální reprezentace. Bez centrálního jádra by konkrétní reprezentace přestala existovat, nebo by změnila svůj charakter.“[13] [Wanger a Valencia 1996: 332] Sociální reprezentace *jádra* jsou organizované, sdílené a udržované sociálními strukturami. Charakterizuje je silné ukotvení, které je v průběhu času neměnné, nebo téměř neměnné. Naproti tomu sociální reprezentace *periferie* vznikají pomocí kognitivních procesů jedince, které se „nabalují“ na reprezentace jádra. Nejsou tolik organizované a

[11] Jejich pojetí rozdělení sociálních reprezentací mi přijde přehlednější než Moscoviciho. Proto uvádím je.

[12] V originále *core* a *periphery*.

[13] V originále: *„One set of these elements, the central core, is the constitutive part of a social representation. Without the central core a specific representation would cease to exist or would change its character.“* [Wanger a Valencia. 1996: 332]

jejich podoba a množství se individuálně liší – do jisté míry mohou být i svým způsobem nahodilé. [Wanger a Valencia 1996: 331–351] Důležité je, že tyto reprezentace jádra jsou proměnné v čase, což může mít za následek proměnu vnímání reality konkrétního jedince. To však Wanger ve své práci již nezmiňuje.

V práci to však zmiňuje Marková a spol. [Marková a spol.: 1998] Ve své práci se zabývá rozdíly v myšlení jedinců v šesti evropských státech (3 západoevropské a 3 post-komunistické). Výsledky jejího zkoumání ukazují, že sociální reprezentace se v čase proměňují a mění tak chování nejen jedince, ale i celé společenské struktury. Marková v zásadě navazuje na Wangera a Valenciana. Sociální reprezentace periferie vznikají pomocí kognitivních procesů jedince a jsou tudíž velmi frekventované. K jejich tvoření, či prohlubování tak dochází prakticky neustále. Po dostatečném prohloubení se ze *sociální reprezentace periferie* stane *sociální reprezentace jádra*. Tento proces funguje samozřejmě i opačným směrem, kdy reprezentace jádra působí na reprezentace periferie. Proměňují se však pomaleji než reprezentace periferie a mají zásadní vliv na formování vědomí nějaké sociální struktury. [Marková a spol. 1998: 805–827] Důležité je však říci, že všechny tyto procesy se odehrávají v sociální struktuře. Sociální strukturu představují myšlenky a pocity sociální skupiny, které se projevují charakteristickým chováním jednotlivých aktérů uvnitř sociální skupiny. Výměna probíhá mezi sociálními reprezentacemi jádra uvnitř struktury a reprezentacemi periferie na okraji struktury. [Marková a spol. 1998: 805–827]

Důležité je však upozornit na silnou asymetrii výměny reprezentací mezi jádrem a periferií. Velké množství reprezentací periferie má vliv na sociální reprezentace jádra, které se následně proměňují. To může mít za následek proměnu sociálních praktik ve společnostech. Marková a spol. tak píšou o dynamičnosti sociálních reprezentací. Marková a spol. upozorňují na důležité propojení mezi sociálními reprezentacemi a sociální strukturou. Reprezentace jsou spojeny s určitým vnímáním reality, které je charakteristické pro nějakou sociální strukturu, a to na základě předchozích socio-historických zkušeností dané struktury. V praxi to znamená, že každá sociální struktura může určitou realitu vnímat jinak. [Marková a spol. 1998: 797–828] V rámci mého výzkumu to znamená, že mí respondenti by teoreticky měli vnímat svého virtuálního asistenta rozdílně, neboť jejich sociální reprezentace jsou determinovány odlišnými sociálními strukturami a jinými individuálními zkušenostmi. Marková a spol. také uvádí, že pro každou sociální strukturu jsou různé sociální reprezentace jinak důležité. Mluví o hierarchii sociálních reprezentací, kdy reprezentace jádra a periferie mohou mít u každé sociální struktury jinou intenzitu. Reprezentace jádra mají silné provázání se strukturou a silně ji ovlivňují. Naopak periferní reprezentace jsou pro strukturu okrajové a nemají na ní takový vliv. [Marková a spol. 1998: 797–828] Určení intenzity a kategorizování sociálních reprezentací však není cílem této práce, proto se jejich hierarchizaci již nebudu dále

věnovat. Sociální reprezentace se neustále proměňují v čase. To má zásadní vliv na proměňování sociální reality a sociální praktiky jednotlivců i struktur. [Marková a spol. 1998: 825] Na základě toho předpokládám, že proměny sociálních reprezentací budou ovlivňovat i proměnu důvěry uživatelů ve virtuální asistenty.

Elcheroth ve své práci vytvořil koncept moderního pojetí teorie sociální reprezentace na základě předchozích zjištění jiných autorů, a vytvořil tak rámec, v jakém je potřeba sociální reprezentace chápat a jak k nim přistupovat. Nejprve zformoval předchozí vnímání sociálních reprezentací autorů do čtyř stručných bodů. [Elcheroth, Doise, Reicher 2011: 730]

1. Znalost světa jedince je sdílena se znalostmi sociální skupiny kontinuálně stejného přesvědčení. Co utváří sociální chování jedince, jsou vlastně sdílené sociální znalosti a znalost věcí používaných na sociální úrovni. To v konečném důsledku ovlivňuje jedincovu mysl v tom, jak takovéto věci vnímá. [Elcheroth, Doise, Reicher 2011: 733 a 737]
2. Sociální reprezentace je nutno chápat jako meta-znalosti. Znalosti a povědomí jedince jsou ovlivňovány tím, co si jedinec myslí, že si myslí jiné skupiny.[14] Jedinec tak reaguje na základě jeho myšlenek o tom, co ví a co si myslí, že ví o myšlení ostatních lidí / ostatních sociálních skupin. [Elcheroth, Doise, Reicher 2011: 733 a 739]
3. Sociální reprezentace se týkají nejenom myšlení jedince, ale i jeho chování. Jedinec reaguje na podněty na základě „uzákonění znalosti“ – tady na základě procesu ukotvení a objektivizace, dle definice Moscovice, která je popsána víše. [Elcheroth, Doise, Reicher 2011: 734 a 742]
4. Sociální reprezentace nejsou pouhým obrazem jedincovy reality, ale zároveň realitu vytváří. Elcheroth uvádí jako příklad: občanství, akademické tituly, třídy, nebo hypotéky, na kterých demonstruje, že tyto pojmy existují na základě toho, že v ně společnost věří. Důležitou vlastností sociální reprezentace je také to, že může v průběhu času měnit svoji povahu a tím i povahu reality jedince, a to na základě nových sociálních reprezentacích, které si jedinec přisvojil. [Elcheroth, Doise, Reicher 2011: 734 a 744]

Tyto čtyři body shrnují poznatky autorů o teorii sociálních reprezentací a poskytují tak jednotný základ pro další pochopení a práci se sociálními reprezentacemi. Elcheroth reaguje svými čtyřmi body. V nich ukazuje, jak je potřeba s reprezentacemi analyticky pracovat, a jak je interpretovat, tak aby byla zachována jejich myšlenka, ale zároveň aby bylo dosaženo cíle a pochopení klíčových problémů. Jedná se vlastně o návod, jak by měl badatel o reprezentacích uvažovat v rámci výzkumu

[14] Pro možnou složitost tohoto bodu zde uvádím i originální znění textu: „…the critical factor in what we do is often less what we think ourselves than what we think others are thinking—both the beliefs of those who stand within our communities and also those who stand against our communities. In this sense, we need to pay as much attention to meta-knowledge (…).” [Elcheroth, Doise, Reicher 2011: 733]

a jak by k nim měl přistupovat.

1. Elcheroth radí, aby se badatel zaměřil spíše na vztahy mezi jednotlivými entitami, či pojmy než na samotného jednotlivce. V opačném případě by badatel potřeboval rozsáhlé informace o celkovém kontextu, nebo o sociální síti jednotlivce. [Elcheroth, Doise, Reicher 2011: 736–738] V kontextu mé práce je tedy potřeba se zaměřit na vztah mezi umělou inteligencí (reprezentovanou inteligentním virtuálním asistentem) a jejím uživatelem. Na základě prostudováním jejich vztahu následně popíšu vzájemné působení technologie a uživatele.
2. Je potřeba se dotazovaných ptát nejenom na to, co si myslí, ale také na to, co si myslí, že si myslí ostatní. Právě dotazy na vnímání myšlení ostatních lidí, či skupin dá badateli odpověď na to, jak dochází k formování reprezentací a jak ovlivňují jedincovo chování. [Elcheroth, Doise, Reicher 2011: 738–740] V mém případě budu ověřovat, zda si uživatelé virtuálních asistentů myslí, že ostatní uživatelé jsou stejného přesvědčení ohledně virtuálních asistentů jako oni samotní a na základě toho se k nim budou i jinak chovat. V neposlední řadě autor také upozorňuje na důležitost meta-reprezentačních výzkumů, které jsou, podle Elcherotha, stejně důležité, jako průzkumy veřejného mínění. To znamená, že by se badatelé měli i zaměřit na zkoumání toho, co si respondenti myslí, že si myslí jejich okolí.
3. Je nutné sledovat kolektivní praktiky. Cílem je zjistit, jak se sociální reprezentace ukládají do kolektivního vědomí. [Elcheroth, Doise, Reicher 2011: 740–743] V rámci mé práce budu zkoumat jaké zkušenosti a prožitky s umělou inteligencí si její uživatelé pamatují nejvíce. Odpovědi mi pomohou při popisování, jakým způsobem se sociální reprezentace ukládají a jaké faktory měly na jejich uložení vliv.
4. V neposlední řadě by se měl badatel zaměřit na pozorování organických celků. Tím je myšleno sledování, jak sociální reprezentace transformují realitu. [Elcheroth, Doise, Reicher 2011: 754–755] Ve své práci tak budu zkoumat, jak prožitky a zkušenosti s používáním virtuálního asistenta ovlivňuje nahlížení uživatele na umělou inteligenci a také jak ovlivňují uživatelovu důvěru v ní.

Možnosti užití sociálních reprezentací

Sociální reprezentace je potřeba brát nejen jako teorii zabývající se chováním sociálních aktérů a způsoby jejich vnímání a interpretace sociální reality, ale také jako nástroj zkoumání. Například Wolfgang Wanger popisuje různá využití teorie sociální reprezentace v empirickém výzkumu. Uvádí, že teorie si klade za cíl překonat dichotomii mezi jednotlivcem a sociální strukturou, stejně tak jako i překonat předěl mezi subjektivitou a objektivitou. Sociální reprezentace tak vytváří mezi

nimi logické spojení, které může badatel zkoumat. Výsledkem je, že pomocí výzkumu sociálních reprezentací dokáže badatel zkoumat makrosociální jevy ve společnosti v jejich historickém a kulturním kontextu. [Wanger a spol. 1999: 100–101] Sociální reprezentace mají rozsáhlé využití napříč všemi metodami výzkumu – od kvalitativních metod po metody kvantitativní. Sám Wanger ve své knize uvádí sedm různých výzkumů, při kterých byly využity sociální reprezentace od etnografické studie, přes mediální analýzu po využití reprezentací při experimentálních návrzích výzkumu.

První příklad, který uvedu, se zabýval etnografickým výzkumem pohlaví. Výzkumu se účastnily pouze děti a výsledem mělo být, jak pohlaví ovlivňuje život dítěte a jak nad ním uvažuje. Pomocí hloubkových výzkumných rozhovorů s použitím přístupu sociálních reprezentací došli badatelé k poznatku, že činnost dítěte je pohlavím silně redukována – dělají, co je ženské, nebo mužské. S tím souvisí i silná polarizace rozdělení věcí na mužské a ženské. V neposlední řadě zjistili různorodost jedinců ve vyjadřování své pohlavní identity. [Lloyd, Dunveen: 1992] Možnosti bádat v celkovém historickém i kulturním kontextu s použitím teorie sociální reprezentace ukáže i druhý příklad. Ten zde uvádím z důvodu propojení kvalitativního a kvantitativního přístupu výzkumu. V rámci zkoumání veřejného a politického života v Brazílii zkombinovali výzkumníci obsahovou analýzu médií a následnou focus group. Opět s využitím teorie sociální reprezentace došli k závěru, že lidé ve veřejném životě jsou historicky silně ovlivněny politickým životem, konkrétně jsou ovlivněny zklamáním, které v politickém životě zažili (například: korupce politiků). Tato zkušenost se následně projevuje do jejich vnímání veřejného života a do jisté míry i součástí jejich kultury. V rámci tohoto výzkumu také autoři potvrdili, že sociální reprezentace jsou neoddělitelné od historického a kulturního kontextu. [Wanger a spol. 1999: 104–106][15]

Teorie samotná je však v podání Moscovice celkem komplikovaná. Přestože jsou její předpoklady a nosné principy autorem jasně vysvětleny (shrnuji v předchozí části *Teorie sociálních reprezentací podle S. Moscovice*), sám autor nikde neuvádí jednoznačnou definici sociálních reprezentací. To dává dostatečný prostor pro námitky a kritiky. Jednou z hlavních námitek teorie její přílišná obecnost a nejednoznačnost definice, což umocňuje abstraktnost už tak dost abstraktnímu pojmu. To může mít za následek nesprávné používání teorie a získání mylných výsledků. [Abric: 1996] V případě mého výzkumu však tato námitka nehraje přílišnou roli, neboť zkoumám obecně fungující princip – tedy to, jak lidé uvažují nad umělou inteligencí a k čemu si ji připodobňují. Z tohoto důvodu je obecnost sociálních reprezentací žádoucí.

[15] Detailní popisy metod a analýz výzkumů, včetně konkrétních kroků využití sociální reprezentacích v praxi popisuje kniha *Theory and Method of Social Representations*, str. 101–121.

Sociální reprezentace jako nástroj pro zkoumání důvěry

V předchozích částech jsem představil různé koncepce pojetí důvěry podle různých autorů, na které jsem následně navázal stručným popsáním teorie sociální reprezentace a jaké možnosti využití teorie ve výzkumu nabízí. Nyní je ještě potřeba čtenáři přiblížit vztah mezi oběma sociologickými koncepty v kontextu této práce.

V kapitole o důvěře jsem psal o různých pojetí důvěry ve společenských strukturách. Pro někoho může být důvěra očekávání určitého chování. [Luhmann 1979: 18–22] Pro jiné lidi je zase důvěra ochotou jít dobrovolně do nějakého vztahu. [Johns 1996: 81] Nebo může být důvěra brána jako postoj. [Rotter 1967: 651–652] Existuje tedy mnoho způsobů, jak důvěru chápat. Většina definic se shoduje v tom, že důvěra je druhem vztahu mezi sociálními aktéry. Proto je k ní potřeba přistupovat v určitém časovém a prostorovém kontextu, ve kterém se možno důvěru zkoumat, popisovat a vysvětlovat. Správné pochopení souvislostí je proto zásadní. Jedním z cílů této práce je objevit zdroje důvěry v umělou inteligenci a zjistit, jakým způsobem je důvěra v umělou inteligenci ovlivňována sociálními, politickými, ekonomickými a kulturními faktory. Abych mohl na tyto otázky odpovědět, je potřeba zvolit správný nástroj zkoumání.

Teorie sociálních reprezentací mi představuje nástroj pro zjištění, jak se důvěra v umělou inteligenci může formovat. Většina výše zmíněných autorů rozděluje reprezentace ovlivňující sociální strukturu na reprezentace jádra a periferie. To může znamenat, že i důvěra se formuje na základě těchto dvou druhů reprezentací. Zdroje důvěry by pak neměli mít všechny stejný charakter, ale měli by mít rozdílný základ podle toho, zda vychází z reprezentací jádra, nebo periferie. O tom například Lewicki a Bunker [Lewicki, Bunker: 1995] vůbec neuvažují. Navíc hierarchičnost reprezentací může způsobit, že i lidé ze stejné sociální struktury (uživatelé inteligentních virtuálních asistentů), budou vnímat umělou inteligenci jinak. Vzájemné působení reprezentací jádra a periferie může dokázat, že definice důvěry v podání Lee a See[16] [Lee, See 2004] nemusí být přesná, neboť do procesu vstupují i jiné faktory závisející nejen na zkušenostech uživatele, ale i na kontextu, vnímání okolí a zažitých stereotypech. Hierarchičnost naopak může být v souladu s formováním důvěry podle Liu, Yang a Xu [Liu, Yang, Xu: 2018], kteří ji rozdělují podle přímých a nepřímých účinků. Teorie sociálních reprezentací mi pomůže odhalit, jak tato dvě pojetí se mohou potvrdit, popřípadě doplnit.
Obecně mnoho autorů zmíněných v předchozí kapitole *Důvěra* uvažuje o jejím formování jen v jedné dimenzi. Výjimkou je Frederiksen [Frederiksen: 2012], který o důvěře uvažuje i rámci času, prostoru i určitého kontextu. Jeho práce je však zaměřená na zkoumání důvěry mezi lidmi, nikoli mezi jedincem a umělou inteligencí. Proto je potřeba určit sociální reprezentace uživatelů virtuálních

[16] „…sociálně psychologický koncept, kdy agent pomůže dosáhnout cíle jednotlivce, v situaci, která je charakterizována nejistotou a zranitelností.“ [Lee, See 2004: 55]

asistentů a porovnat je s Frederiksenovými dimenzemi.

Dle teorie sociálních reprezentací je nutné se na důvěru dívat optikou zaměřenou na i sociální a kulturní myšlení jednotlivců. Díky tomu mohu sledovat, jak jsou nová sociální poznání a reprezentace reality vytvářeny interakcemi ve společnosti a jaký vliv na tento proces mají okolní faktory. Komplexnost teorie umožňuje zkoumání sociálních reprezentací na různých úrovních (jednotlivců i struktur) a v kontextu současného dění. V neposlední řadě jsem se pro využití teorie sociálních reprezentací rozhodl, kvůli možnosti úzce propojovat teoretickou a empirickou úroveň zkoumaného problému. [Höijer 2011: 3–16]

Teorie sociálních reprezentací mi jako nástroj poskytuje návod, jak důvěru uživatelů v inteligentní virtuální asistenty zkoumat. V první řadě je potřeba se zaměřit na vztahy. Jedním ze zkoumaných vztahů je vztah mezi uživatelem a jeho virtuálním asistentem, který využívá v mobilním zařízení, v počítači, nebo na tabletu. Stejně tak je i potřeba zjistit, zda existuje nějaký vztah mezi jednotlivými uživateli virtuálních asistentů a popřípadě, jak tento vztah ovlivňuje uživatelovu důvěru v samotného asistenta. Dále je také potřeba se zaměřit na to, jak uživatel pracuje s informacemi o inteligentních virtuálních asistentech. Tím je myšleno, odkud uživatel sbírá informace o virtuálních asistentech, jakou váhu informacím přikládá a jak s nimi nadále pracuje. V neposlední řadě se zaměřím, jak informace o virtuálních asistentech ovlivňují uživatelovo vnímání reality, potažmo, jak různé zkušenosti transformují uživatelovu realitu. Takto stručně řečeno mi teorie sociální reprezentace umožňuje vnímat důvěru v širších souvislostech a stává se tak z něj ideální nástroj pro můj výzkum.

Sociologická reflexe technologického vývoje

Nové technologické možnosti jako proměna virtuálního prostředí World Wide Webu, větší výpočetní kapacita domácích počítačů, notebooků, či tabletů, nebo prostě nové funkce technologií mají za důsledek proměnu chování některých jejich uživatelů. Ústředním motivem procesu propojování jsou tzv. *nová média*. Novými médii je myšlen Internet, počítače, mobilní telefony, či sociální sítě. Tedy entity, které se staly základními atributy každodenního života jednotlivců v mnoha společnostech. Důležitý je jejich vliv na transformaci sociální reality uživatelů. Jejich využíváním dochází k propojování a kombinování reality a virtuality, čím dochází k transformování uživatelova vnímání sociální reality. [Kyslova, Berdnyk 2014: 68–70] Zkoumání nových médií je z pohledu, jednotlivce i sociálních struktur pro sociologii velmi důležité.

Je však potřeba si uvědomit, že důležitost nových médií vychází z možností, které technologii dávají programy s implementovanými algoritmy umělé inteligence. Díky těmto algoritmům dochází ke

kompletní transformaci kybernetického prostoru a tím i dochází k proměně chování uživatelů v něm. V tomto kontextu lze hovořit o počátku tzv. Webu 3.0, pro který jsou nová média klíčová. [Kyslova, Berdnyk 2014: 83–86] Jak již název napovídá Web 3.0 je třetí teoretické rozdělení vývoje samotného World Wide Webu. V případě první fáze Webu, tedy v případě webu 1.0 se mluví o samotných začátcích a budování celého internetu. [The Third-Generation Web is Coming: 2006] Druhá fáze webu – web 2.0 je spojován s možností ukládání či spravování dat prostřednictvím různých zařízení, nikoliv pomocí jednoho konkrétního. S tím je také spojeno rozšíření komunikačního prostředí a jeho masovější využívání. Typickým příkladem je vlastnění a využívání e-mailové adresy. [Lash 2006: 571–581][17] Web 3.0 je také nazýván Intelligent World Wide Web. Vliv nových médií je pro tento druh webu klíčovým. Hlavním atributem webu 3.0 je dominance inteligentního prostředí, tj. technologické prostředí s implementovanými prvky umělé inteligence. Inteligentní prostředí v každodenním životě, přičemž není myšleno pouze prostředí kybernetické, ale jeho vliv se promítá i ve světě reálném (výtahy, chatboti). [Kyslova, Berdnyk 2014: 72–100] Projevy webu 3.0 v reálném světě je možné spatřovat nejen v promítání virtuality v sociálním prostředí jednotlivce, ale i celkovým podílem inteligentních technologií v každodenním životě a jejich potřebou. V obrovském množství informací by bylo pro běžného uživatele velmi obtížné najít na internetu přesně tu informaci, kterou uživatel hledá. Vzhledem k tomu, že většina předních prohlížečů (například Google) má implementovanou schopnost strojového učení, je vyhledávání na internetu pro uživatele po krátké chvíli jednoduché. Takovéto sémantické weby však nejsou jediným příkladem webu 3.0. Dalším jeho atributem je již zmíněné strojové učení, nebo možnost vyhledávání v přirozeném jazyce. Vyhledávače také používají tzv. recommendation agents[18], data-mining a jiné formy implementované umělé inteligence. [The Third-Generation Web is Coming: 2006] Interaktivní weby 3.0 takovým nástrojem jsou pro uživatele atraktivnější a pohodlnější. Uživatel si na ně brzy zvykne, využívá je a začne jim věřit. [Merrilees, Fry 2003: 455 – 472] Podrobněji důvěru v technologie rozpracovávám v části: *Důvěra v automatizaci a umělou inteligenci*.

Toto samozřejmě není taxativní výčet všech možností, které uživatelům přináší nové technologie v podobě nových médií, nebo v podobě schopností webu 3.0. Mým cílem v tomto oddíle bylo ukázat, jaký vliv mají nová média na sociální vnímání reality uživatelů a také ukázat na zvyšující se podíl nových médiích ve společnostech. Větší podíl a lepší dostupnost nových médií má tak vliv i na

[17] Dalšími důležitými prvky webu 2.0 je také možnost ukládání dat do virtuálních uložišť, či možnosti vzdáleného přístupu, které se staly důležitou součástí práce na počítači. Snadná dostupnost a přenášení každodenních věcí do kybernetického prostoru zapříčinilo, že stolní počítače a notebooky se staly běžnou součástí všedního života. Umět pracovat s počítačem se tak začala stávat, důležitou dovedností ve společnostech s rozšířeným webem 2.0. [Beer, Burrows 2007: 1–13

[18] Schopnost programu doporučovat uživatelům obsah na základě jejich předchozího výběru.

proměnu virtuálního prostoru (webu), který využívá nových médií za účelem „přiblížení se“ k uživateli.

Inteligentní virtuální asistent (IVA)

Jak už z úvodu i samotného názvu práce vyplývá, zaměřím se na uživatele inteligentních virtuálních asistentů (IVA). Obliba virtuálních asistentů neustále roste, což je zapříčiněno vyšším podílem chytrých telefonů ve světě a s tím spojený vyšší počet programů a aplikací s implementovanou umělou inteligencí. [Kim, Boelling, Haesler, Bailenson, Bruder, Welch 2018: 105–106]

Vědecké zkoumání vtahu mezi člověkem a IVA navazuje na práce z 90. let minulého století, které se zabývaly vztahem mezi člověkem a počítačem a důsledky jejich vzájemného působení. Výsledky těchto prací přinesly například důkaz, že člověk má tendenci komunikovat s inteligentními přístroji jako s člověkem. Samozřejmě za předpokladu, že stroj vykazuje určité známky inteligence. [Nass a spol. 1995: 228–229] Vnímání technologie tak závisí na kontextu. Právě v jakém kontextu uživatel technologii pozná, tj. v jakém ohledu je o ní mluveno, nebo v jakém prostřední se jedinec nalézá, má významný vliv na to, jak jedinec technologii hodnotí [Blascovich 2002: 133–136] To se samozřejmě netýká pouze virtuálních asistentů, ale všech technologií, které uživateli vykazují známky inteligence. Pozdější výzkumy poté prokázaly na vztah mezi vlastnostmi a vzhledem IVA a lidským uživatelem. Čím více měl IVA lidské vlastnosti a podobu, tím více mu byli uživatelé ochotni důvěřovat. [Kim, Boelling, Haesler, Bailenson, Bruder, Welch 2018: 105–119]

Inteligentní virtuální asistenty zkoumali také Purwanto, Kuswandi a Fatmah, kteří určili jako hlavní výhodu IVA „obousměrnost“. [Purwanto, Kuswandi a Fatmah: 2020] Pod pojmem obousměrnost si může čtenář představit vzájemnou aktivní interakci, kdy jeden z aktérů reaguje na toho druhého, tedy kdy IVA reaguje na uživatelovi rozkazy a opačně. IVA jsou oblíbené kvůli třem důležitým vlastnostem. V první řadě se jedná o *ovladatelnost*. Inteligentní asistenty je možno ovládat hlasovými povely.[19] Není tak potřeba cokoliv psát, či držet telefon přímo v ruce, ale je možné ho ovládat pohodlně na dálku. Uživatelé tak na něm mohou oceňovat jeho schopnost myslet a plnit požadavky. Druhou důležitou vlastností je *synchronicita*. Rychlost, kvalita a spolehlivost jsou důležité požadavky uživatelů IVA, které je provádí a může je provádět i prostřednictvím jiných platforem či technologií. Přes virtuální asistenty je tak možné dnes ovládat například televizi, ledničku, či celou domácnosti (např. Alexa od firmy Amazon). Třetím důležitou vlastností je *vzájemnost*. Vzájemnou komunikací a interakcí se oba aktéři učí. Uživatel se učí lépe využívat svého virtuálního asistenta,

[19] Z pravidla jsou povely potřeba říci anglicky, ale pro některé virtuální asistenty si již spouští česká verze [applenovinky.cz: 2019]

asistent postupem času lépe chápe povely „svého pána“ na základě jeho předchozích požadavků. Autoři však upozorňují na riziko zneužití osobních údajů, které si uživatelé málo uvědomují (viz předchozí část). Na druhou stranu, právě riziko ztráty osobních údajů vnímají uživatelé IVA jako to největší riziko, které se jim může při interakci s asistentem stát. [Purwanto, Kuswandi a Fatmah 2020: 64 – 75] Jejich práce však už nezkoumá, kdo za potenciální ztrátu osobních údajů může, respektive, jakou roli v něm hraje inteligentní virtuální asistent. I touto otázkou se ve své práci hodlám zabývat.

Schopnosti inteligentních virtuálních asistentů jsou rozsáhlé. V současnosti je možno s takovým asistentem jako je například Siri, nebo Axela mluvit a jako s obyčejným člověkem a žádat je o splnění určitých úkonů. Je možné se IVA zeptat: „Co mám dnes v kalendáři za události?“ Kdo má dnes narozeniny? Za jak dlouho přijedu do práce? Nebo objednej mi jídlo. V případě inteligentních domácností je také možno se ptát: Co mám v ledničce? Co mám koupit? V 17:55 zapni troubu na 150 stupňů. Nebo ve 01:00 vypni televizi. To však sebou nese také nebezpečí zneužití nebo chyby asistenta. Při práci s IVA může dojít k zcizení, upravení, zneužití, či jiného neoprávněného zacházení s daty uživatele třetí osobou nebo programem. Také je potřeba počítat s chybovostí samotného asistenta, ke kterému může dojít. [Chung, Iorga, Voas, Lee 2017: 100–104]

Ani v tomto případě nelze uvést jednotnou definici inteligentního virtuálního asistenta. Důvodem je, že tito asistenti vychází z technologické tradice vývoje tzv. chatbotů, tedy strojů, které dokáží vést s uživatelem věcný dialog na určité téma[20], přičemž by uživatel neměl poznat, že si píše se strojem. Jedná se tak o odpověď vývojářů na Turingův test, který spočívá právě v tom, že uživatel nesmí poznat, kdy komunikuje se strojem a kdy s jiným člověkem – tak se podle Alana Turinga pozná umělá inteligence. [French 1990: 53–65] První chatbot ELIZA byl vytvořen Josephem Weizenbaumem v 50. letech minulého století. Za inteligentního virtuálního asistenta jsou dnes považovány aplikace, či programy založené na principu strojového učení, které usnadňují uživatelům práci v kybernetickém prostoru, a to pomocí příkazů daných uživatelem. IVA může mít pojmenování také například: inteligentní asistent, inteligentní osobní asistent, digitální asistent, či osobní virtuální asistent. [Chung, Iorga, Voas, Lee 2017: 101] Mezi nejznámější virtuální asistenty jsou v současnosti Siri (Apple), Bixby (Samsung), Google Assistant (Google), M (Facebook), Cortana (Microsoft), Alexa (Amazon), nebo Alisa (Yandex).

Nabízí se nyní otázka, proč jsem zvolil IVA jako zástupce umělé inteligence, když i samotní virtuální asistenti o sobě tvrdí, že nejsou klasickým příkladem umělé inteligence, nebo odpovídají

[20] Záleží k čemu chatbot určený

vyhýbavě.

Inteligentní virtuální asistenti mají v sobě implementované schopnosti strojového učení. „Bez ohledu na terminologii ‚mozek' systému – inteligence, která převádí lidský hlas na text, provádí lingvistickou analýzu, na jejímž základě provádí požadovanou akci…" [Chung, Iorga, Voas, Lee 2017: 101] Jedná se tedy opět o transformaci dat, jejímž výstupem je účelné rozhodnutí – proces automatizace. [Lee, See 2004: 50] V případě slabého pojetí umělé inteligence, tak jak ji prezentují Russell a Norvig, záleží na tom, zda umělá inteligence dokáže konat specifické úkoly, které vyžadují lidské schopnosti. [Russell, Norwig 2010: 1–17] Inteligentní virtuální asistenti tedy dokáží přijímat informace a na jejich základě provést účelnou adekvátní reakci. Na základě výše zmíněných definic mohu IVA považovat za umělou inteligenci.

Uvědomování si rizika

V předchozí části jsem uvedl příklady využití nových médií ve společnostech a taky jsem popsal důležitost strojového učení, potažmo umělé inteligence pro funkčnost moderních webů generace 3.0. V následující části popíši, jak nové možnosti webu a moderních médií obecně působí na lidské chování. Z předchozí části jasně vyplývá, že moderní technologie ovlivňují lidské chování na denní bázi a není mým cílem popisovat veškeré vlivy technologií na jedincovo uvažování. Proto se nyní zaměřím pouze na uvažování / chování jedince ve vztahu k uvědomování si rizik spojených s využíváním technologie. Důvodem je silné propojení mezi důvěrou v něco a uvědomováním si rizika, což je také důležitou částí této práce.

Výše jsem uváděl, že podle Lee a See [Lee, See: 2004] nemá umělá inteligence vlastní úmysl, tedy, že lidé mají tendenci ji kvůli tomu věřit. Uvědomují si tedy lidé nějaká rizika při práci s virtuálními asistenty, nebo jim pouze slepě důvěřují? Velmi proklamovaným rizikem je zneužití osobních údajů, které si u mediálně často řeší. [Belanger, Xu 2015: 573] Na druhou stranu řada technologií po uživatelích žádá osobní údaje, přičemž jim většina uživatelů vyhoví. Většina současných chytrých telefonů má v sobě již zakomponovanou GPS, akcelerometry a další prvky umožňující sledování. Uživatel se do telefonu musí zaregistrovat pomocí e-mailové adresy. Některé telefony vyžadují i zadání čísla platební karty. [Lupton 2014b: 4–5] Vývojáři softwaru, prodejci, či zájmové skupiny – všichni tito aktéři mohou monitorovat uživatelovu aktivitu v kybernetickém i reálném prostoru a to za účelem zkvalitňování jejich produktů. Luptonová píše o tzv. self-trackingu, tedy o procesu, kdy uživatel, vědomě, či nevědomě monitoruje své chování pomocí technologie a data následně odesílá do virtuálního prostředí. Jedná se o tzv. sefl-trackingové kultury, kdy jednotlivci využívají výhod moderních technologií, aby získaly více informací sami o sobě, či o dané skupině. [Lupton 2014a: 77 – 86] Nejzřetelněji je zvyšující se obliba vidět v oblasti zdravotnictví a veřejného zdraví. Autorka

v tomto kontextu mluví o sebe-zodpovědnosti. Na uživatele se přenáší osobní zodpovědnost za jejich tělo a za jejich zdraví. [Lupton 2014a: 77 – 86]

Je však potřeba si položit otázku, do jaké míry lze hovořit o dobrovolnosti využívání. Je pravdou, že za normálních okolností člověka nikdo nenutí podobné technologie využívat. Na druhou stranu je člověk přímo, či nepřímo manipulován k využívání monitorovacích prostředků – povolování souborů cookies, které monitorují činnost jedince na internetu. V případě, že uživatel webu nesouhlasí, nezobrazí se mu celý obsah stránky. Nebo sledování zdravotní stavu prostřednictvím aplikace. Některé zdravotní pojišťovny nabízí slevu na pojištění, při dodržování zdravého životního stylu. K monitorování je však potřeba aplikace, které monitoruje činnosti jednice (například počet uběhnutých kilometrů). Kromě údajů o zdraví jednice (dle kategorizace GDPR spadají do zvláštní skupiny citlivých údajů [orientace v GDPR: 2020]) podobné aplikace zpracovávají i další osobní údaje uživatelů, které však uživatel nemá šanci kontrolovat.

Právě nedostatečná kontrola zpracovávání osobních údajů může být ve vztahu k novým technologiím potencionální problém. [Belanger, Xu 2015: 573] Vznikají zde dvě potencionální rizika Buďto riziko existence nedostatečně rozvinutého regulačního aparátu, který nedokáže uživatele nových médií adekvátně chránit před nebezpečím a zneužitím osobních údajů, anebo je zde riziko naopak přílišného regulačního rámce, který výrazně redukuje možnosti uživatelů zcela svobodně užívat nová média.

Jiným druhem rizika se zabývá Mirilello a spol. Upozorňují na zvyšující nerovnost mezi komunikačními možnostmi a lidským časem. Díky moderním technologiím nemusí být vzdálenost překážkou v komunikaci. Lidé mohou komunikovat prakticky neustále, což prohlubuje vzájemný vztah mezi jedinci. Na druhou stranu čas, který mohou lidé komunikaci obětovat, je omezený. Lidé sice pomocí komunikačních technologií mohou nyní komunikovat častěji a intenzivněji než dříve, mohou však obsáhnout mnohem menší sociální prostor. Člověk rozděluje svůj komunikační čas mezi menší skupinu lidí a tím si oslabuje své vazby na širší sociální okolí. Při častém využívání komunikačních technologií dochází tvorbě vazby mezi uživatele a jeho komunikačním prostředkem. Uživatel si na něm vytváří vazbu a určitou závislost. Mirilello uvádí, že si uživatelé mohou se svým mobilním telefonem vybudovat vztah, neboť s ním tráví hodně času. Stává se pro ně důležitým, nechtějí ho nikomu svěřovat a jsou a něj velmi opatrní. [Mirilello a spol. 2013: 89–95]

Dalším potencionálním rizikem může být ztráta lidské dominance. Přesto, že úroveň schopností umělé inteligence není v současné době na takové úrovni, aby mohla konkurovat ve všech ohledech lidským schopnostem, existují obavy z jejího potencionálu. Na potencionální nebezpečí ze strany

schopností umělé inteligence často upozorňuje popkultura sci-fi. [Boström: 2017] „Tradice sci-fi je původně literární žánr, nazývaný též vědecko-fantastickou literaturou, vycházející nejenom z představ o světě, ale i zkušeností o jeho fungování, inspirovaný pokrokem ve vědě a technice. (…)Často je průmětem obav z hypertrofie technické dimenze civilizace. Může mít také podobu varovné prognózy i sebenaplňujícího se proroctví.“ [Maříková, Petrusek, Vodáková a spol. 1996: 969–970]. Člověk je hierarchicky nadřazeným druhem a o tuto výsadu nechce přijít. Peggs tuto hierarchii dokazuje vzájemnou trvalou nerovnost mezi domácími mazlíčky a jejich pány. Přestože je v mnoha případech pes, či kočka brána jako člen domácnosti, nikdy nebude mít stejná práva, jako lidé. Dále ukazuje, že i přístup sociologického bádání je v tomto smysle nevyrovnaný a silně protěžující lidskou populaci. [Peggs 2013: 591–606]

Uvědomování si rizika také závisí na pozici jedince ve vztahu k technologii. Nelze očekávat, že stejný názor na umělou inteligenci bude mít její běžný uživatel, její distributor, výrobce, či člověk, který kvůli automatizaci přišel o zaměstnání. Je proto potřeba při výzkumu uvažovat o vzájemném vztahu mezi jedincem a technologií. Vzhledem k jejich rozdílným pohledům na technologie mají následně tendenci se mezi sebou vymezovat. Sun a Medaglia to ukazují na příkladu, kdy se lékaři v jedné čínské nemocnici silně vymezovali proti diagnostickému robotovi, kterému nedůvěřovali a nedávali mu složité úkoly. Naopak výrobce technologie prosazuje jeho hlubší integraci do zdravotního systému. Dalším aktérem byla vláda, která ovšem také na pozici umělé inteligence ve zdravotnictví měla svůj názor, hájící zájmy vlády, který nebyl v souladu s názory lékařů, ani výrobců technologie. [Sun, Medaglia 2019: 368-383]

Madhavan a Wiegmann uvádí, že lidé se orientují na základě povrchových charakteristik o dané technologii. Pokud je tedy popis nějaké technologie založen na její odbornosti a důvěryhodnosti, lidé to dále nezkoumají. [Madhavan, Wiegmann 2007: 285] K tomu je ještě potřeba doplnit důležitost faktoru značky. Lidé důvěřují silným značkám. Pokud není jedinec příliš limitován finančními prostředky, při výběru technologie většinou sáhne po produktu od známé a prověřené značky. Právě značka je pro uživatele „garantem“ spolehlivosti a kvality a to i v případě, že se jedná o zcela nový produkt. To samé platí i u technologií využívající inteligentní rozhraní, či strojové učení. V případě, že se jedná o produkt ekonomicky silné, či známé značky (Apple, Google, Amazon, Samsung), jsou jejich aplikace a rozhraní všeobecně lépe přijímaná než stejná rozhraní od menších, či méně známých firem, protože jsou již osvědčené.[21] [Brill, Munoz, Miller: 2019]

[21] Autoři ve svém textu neuvádí další možné příčiny lepšího přijímání technologie uživateli jako je lepší marketing, vetší prodej, širší dostupnost a podobně. Dá se ovšem logicky vyvodit, že veškeré tyto možné příčiny mají svůj původ buď právě ve značce, kdy si společnost může do marketingu a prodeje dovolit investovat více peněz, anebo naopak právě tyto faktory určují, která značka je ekonomicky silná, či známá.

Důvěra, sociální reprezentace a virtuální asistenti

V této kapitole popíši, jak spolu souvisí koncept důvěry, teorie sociálních reprezentací a inteligentních virtuálních asistentů v kontextu mé práce. Hlavním předpokladem je, že důvěra jedince v umělou inteligenci má svá specifika oproti důvěře mezi dvěma jedinci. V kapitole *Důvěra* jsem psal o různých hlediscích, jak je možné důvěru vnímat (např.: naplnění očekávání, předmět důvěrného vztahu, vzájemný vztah zúčastněných stran). Většina autorů se na důvěru v technologie dívalo obecně bez širšího kontextu pouze z jednoho z těchto hledisek [Lee, See: 2004; Madhavan, Wiegmann 2007; Han, Sang-Lin 2007; Boyd, Crawford 2012]. Tento přístup však nemusí být dostačující.

V předcházející kapitole jsem psal o výhodách inteligentních virtuálních asistentů. Purwanto, Kuswandi a Fatmah určili jako hlavní výhodu IVA *obousměrnost*, tedy schopnost aktivní interakce. [Purwanto, Kuswandi a Fatmah 2020: 64 – 75] To naznačuje, že by se k virtuálním asistentům mohlo ve vztahu k důvěře přistupovat podobně, jako v případě interakce dvou a více jedinců. Jejich další závěry ale definují tři důležité vlastnosti IVA, kvůli kterým jsou tolik oblíbení, jsou to: *ovladatelnost*, *synchronicita* a *vzájemnost*. [Purwanto, Kuswandi a Fatmah 2020: 64 – 75] Pokud jsou toto pro uživatele skutečně důležité vlastnosti, nelze přistupoval k vzájemné důvěře stejně, jako ve vztahu dvou jedinců.[22] Ovladatelnost, synchronicita i vzájemnost jsou dimenze, kterými se autoři, zabývající se důvěrou v technologie a umělou inteligenci, vůbec nezabývají. Nelze ani vyloučit, že existují další úhly pohledů, ze kterých je potřeba důvěru v umělou inteligenci zkoumat.

Z tohoto důvodu jsem se rozhodl pro zkoumání důvěry v inteligentní virtuální asistenty, potažmo pro zkoumání důvěry v umělou inteligenci prostřednictvím teorie sociálních reprezentací. Teorie mi poskytuje nástroj pro určení důležitých aspektů, které si uživatelé ve vztahu k IVA vybavují. Především je potřeba zjistit, jak svého virtuálního asistenta vnímají a na základě toho určím, jaké sociální reprezentace mají se svým IVA spojené. Různé druhy sociální reprezentací (jádra a periferie) naznačují, že ne všechny zkušenosti (přímé, či zprostředkované) budou mít na uživatele stejný vliv. Znamená to tedy, že některé zdroje důvěry mohou být stále, zatímco nějaké mohou být v čase proměnlivé. Teorie sociálních reprezentací mi pomůže zjistit, do jaký míry se dá v tomto ohledu použít temporální rozdělení důvěry podle Lewicki a Bunker [Lewicki, Bunker: 1995], když se zabývali především vztahu dvou a více jedinců.

[22] Ovladatelnost a synchronicita by v případě mezilidské interakce vztažené k důvěře postrádala svůj smysl.

Metodologická část

Design výzkumu

Cílem výzkumu je určení hlavních zdrojů důvěry uživatelů virtuálních asistentů a vytvoření typologie, podle které se dá uživatelská důvěra kategorizovat. Dalším cílem je popsání role sociálních reprezentací s ohledem na důvěru a zjištění, zda je důvěra podmíněná sociálními reprezentacemi. Mým cílem tedy je pochopit podstatu vztahů, které se utvářejí mezi lidmi a stoji/programy obsahující umělou inteligenci –> IVA. Teprve pochopením vztahů mezi uživatelem a technologií mohu nalézt klíčové faktory, které ovlivňují motivaci lidí důvěřovat, či nedůvěřovat umělé inteligenci implementované do všedního života a tím zjistit i samotné zdroje důvěry.

Pro pochopení určitého vztahu a porozumění dané problematiky je vhodné využít kvalitativní metodu zkoumání. „Kvalitativní metody se užívají k odhalení a pochopení toho, co je podstatou jevů, o nichž toho ještě moc nevíme. Mohou být také použity k získání nových a neotřelých názorů na jevy, o nichž už něco víme. V neposlední řadě mohou kvalitativní metody pomoci získat o jevu detailní informace, které se kvantitativními metodami obtížně podchycují.“ [Strauss, Corbin 1999: 11] Právě potřeba podrobných informací k pochopení vztahu mezi člověkem a technologií je důvodem, proč jsem se rozhodl pro tuto metodu zkoumání. „Kvalitativní metoda výzkumu umožňuje hlubší vhled do určité situace či problematiky za účelem zjištění její podstaty.“ [Strauss, Corbin 1999: 10] Tato metoda poskytuje mnohé účinné nástroje na získání komplexního pohledu na celkovou oblast, včetně určitého kontextu, a navíc mi dává možnost nezkoumat pouze povrchové informace, ale mohu se snadněji dostat až k jádru věci. Proto jsem se rozhodl pro svůj výzkum využít polo-strukturované výzkumné rozhovory.

Jak jsem již představil výše, pojem umělá inteligence je poměrně široký a občas i těžko uchopitelný. Proto je potřeba, aby moji respondenti měli jasně danou podobu umělé inteligence –> IVA a teprve poté zkoumat, jaké odlišnosti se projevují v jejich vnímání umělé inteligence v jasně dané podobě. Proto jsem se rozhodl při výběru respondentů použít filtr. Filtrem je v této práci myšlen krátký elektronický dotazník, který jsem rozeslal přes sociální sítě. Tento dotazník slouží k tomu, abych mezi stovkami potenciálních respondentů, vybral takový vzorek lidí, kteří již mají zkušenosti s virtuálním asistentem, umějí s ním zacházet a využívají jeho na rozličné činnosti. Jinými slovy se mi do výzkumu dostanou pouze lidé, kteří svého virtuálního asistenta často využívají a zadávají mu různorodé úkoly. Těchto lidí se následně budu v rozhovorech ptát, kde se vzala motivace důvěřovat umělé inteligenci a jak na ně umělá inteligence působí.

Výzkumné cíle

Hlavním výzkumným cílem, který si klade tato práce za cíl, je: Určení hlavních zdrojů důvěry v umělou inteligenci. S tím je i spojeno určení klíčových faktorů, které tyto hlavní zdroje důvěry ovlivňují. Na základě těchto výzkumných cílů, jsem následně zformuloval tři výzkumné otázky, s jejichž pomocí dokáži své cíle naplnit.

První otázka si klade za cíl, určit hlavní zdroje důvěry participantů v umělou inteligenci, která je implementována do každodenního života. V případě samotných zdrojů o nich uvažuji, jako o zdrojích, které budují důvěru, ale i jako o zdrojích, které budují nedůvěru.[23]

- **Jaké jsou hlavní zdroje důvěry v umělou inteligenci implementované do každodenního života?**

S první otázkou je ještě spojená podotázka, která se zaměřuje na vliv sociálních reprezentací v procesu tvorby důvěry uživatele k technologii. Tato podotázka zní:

- **Jak je důvěra podmíněná sociálními reprezentacemi o umělé inteligenci?**

Druhá otázka je spojená přímo s teoretickým pozadím na určení přítomnosti důvěry, či nedůvěry. A zároveň také prohlubuje vhled do vzájemného vztahu mezi participantem a umělou inteligencí. Výzkumná otázka tedy zní:

- **Jaká rizika si jedinec uvědomuje ve vztahu k umělé inteligenci?**

Poslední otázka se odvíjí přímo od teorie sociální reprezentace. Pomocí této otázky doplním ono „bílé místo“ v sociologickém poznáním a zjistím, jak byla důvěra v umělou inteligenci vybudována a na jakých důležitých pilířích stojí.

- **Jakým způsobem je důvěra v umělou inteligenci ovlivněna sociálními, politickými, ekonomickými a kulturními faktory?**

Průběh výzkumu

Hluboký vhled do problematiky a zároveň pochopení vztahu mezi jedincem a věcí s implementovanou umělou inteligencí vyžaduje rozsáhlejší kvalitativní analýzu, rozdělenou do několika kroků a přesné zacílení a pojmenování problému. V této kapitole tedy popíši, na jaké jednotlivé kroky jsem svůj výzkum rozdělil a jak jsem postupoval.

V první řadě bylo potřeba na základě dostupné literatury identifikovat klíčové oblasti, o které by se měl výzkum zajímat. Na základě prostudované literatury jsem zjistil, že vědecké práce věnují velmi

[23] Uvědomování si rizika v případě interakce s nějakou entitou, může být podmínkou důvěry v pojetí různých sociologických teoretiků. Například Luhmana a jeho pojmy *Trust* a *Confidence*, které jsou přímo závislé na uvažování o rizicích. [Luhmann 1979, 18–22]

malý prostor multidimenzionálním zdrojům důvěry v umělou inteligenci (rozdílnost počátků důvěry, druhům důvěry, vliv temporality, vliv kontextu a prostředí). Podle toho jsem si následně definoval výzkumné otázky uvedené výše, a podle nich následně vytvořil design výzkumu a scénář výzkumných rozhovorů.

Ve druhé fázi jsem přešel k výběru vhodných respondentů. Vytvořil jsem krátký on-line dotazník, který zjišťoval, zda respondent vlastní IVA, zda ho používá alespoň 2x denně po dobu minimálně 3 měsíců. Další podmínkou byla rozličnost činností prováděných prostřednictvím IVA a alespoň základní obecná znalost virtuálních asistentů, tj. vědět k čemu asistent slouží, či jaké jsou nějaké možnosti jeho využívání. Dotazník jsem dal na dvě diskuzní fóra o virtuálních asistentech (letemsvetemapplem.eu[24]; zive.cz[25]) a také jsem ho rozeslal přes sociální síť Facebook. Všechny takto kontaktované respondenty jsem poprosil o součinnost a o poslání dotazníku dalším jejich přátelům, aby se metodou nabalování šířil dál a získával nové respondenty. Dotazník byl navržen velmi jednoduše z toho důvodu, že slouží čistě jako filtr potencionálních participantů a s jeho daty neplánuji dále pracovat.[26] Navíc šetří i respondentův čas. Celkem jsem získal odpověď od 128 lidí, z toho 42 % bylo vyplněno na základě přímého odkazu, který jsem respondentovi zaslal. Zbytek respondentů byl zprostředkovaný. Průměrné vyplnění jednoho dotazníku trvalo 5 minut. Z 324 lidí splňovalo podmínky účasti na výzkumu pouze 19 lidí (využívání IVA a rozličnost činností prováděných prostřednictvím IVA byly nejčastější překážkou pro účast ne výzkumu). Z 19 vhodných respondentů jsem na základě úsudku vybral 10 lidí, kteří nejvíce odpovídali kritériím účasti ve výzkumu, tedy aktivní a dlouhodobé využívání virtuálního asistenta, provádění prostřednictvím asistenta rozličné činnosti a povědomí o virtuálních asistentech obecně. Vyhodnocování vyplněných dotazníků probíhalo kontinuálně, a to až do doby, kdy jsem získal dostatečný počet 10 participantů, jejichž povědomí a znalosti o IVA odpovídají cílům této práce a zároveň jsou ochotni se mou provést výzkumný rozhovor na toto téma.

Ve třetím kroku začal sběr dat z výzkumných rozhovorů. Výzkumné rozhovory měly charakter polo-strukturovaných standardizovaných rozhovorů, tj. otázky byly předem dány a měly otevřený charakter, nicméně byl zde prostor pro doplňující otázky mimo scénář výzkumného rozhovoru. Otázky vycházely z teoretické části, kdy byly definovány základní pojmy a dimenze. Všichni

[24] https://www.letemsvetemapplem.eu/ - diskuzní fórum zaměřené na produkty společnosti Apple.

[25] https://www.zive.cz/ - server zabývající se obecně novými technologiemi.

[26] Dotazník byl vytvořen přes internetový portál www.survio.cz. Každý z dotazovaných obdržel odkaz, který rozkliknul a objevil se mu dotazník. Respondenti měli na zodpovězení dotazníku neomezený čas, mohli se k odpovědím vracet a zpětně je změnit. Nicméně po odeslání vyplněného dotazníku již nemohl být dotazník na základě stejného odkazu opět vyplněn.

respondenti odpovídali na stejné otázky, které měly přímou návaznost na zkoumanou problematiku. Jejich výhodou oproti nestandardizovaným druhům rozhovorů je, že se výrazně redukuje vliv tazatele na vývoj rozhovoru a tím i případné ovlivnění odpovědí respondenta. Navíc také usnadňují samotnou analýzu. Tazatel však může otázky lehce při rozhovoru poupravit, aby v některých případech neztrácely svoji věcnost a relevanci [Hendl a Remr, 2017, 84]. Metodika polostrukturovaných výzkumných rozhovorů umožňuje tazateli položit i doplňující otázky, díky kterým tazatel může mírně korigovat míru obecnosti odpovědí, na jaké se participanti pohybují. Součástí dotazování byla i technika spontánních asociací. Participanti byli dotazováni na první asociaci, která se jim vybaví, například při slovech inteligentní virtuální asistent, Siri, nebo umělá inteligence. Tato technika je vhodná především kvůli vyvolání bezprostředních asociací participantů, které jsou ve své podstatě zásadní pro odhalování zdrojů pocitů lidí ohledně nějaké záležitosti. [Hollway, Jefferson 2008: 296–315]

Výzkumné rozhovory trvaly v rozmezí 45–90 minut. Jejich délka se primárně odvíjela od ochoty participanta o daném tématu mluvit. Na začátku rozhovoru byly participanti seznámeni s účelem a cíli výzkumu. S participanty byla také dohodnuta změna jejich jména ve výzkumu, aby byla zaručena jejich anonymizace, což respondenti stvrdili podpisem informovaného souhlasu. Odpovědi participantů jsem zaznamenával na diktafon, díky kterému jsem mohl aktivně reagovat na průběh rozhovoru a popřípadě si zapisovat poznámky / zajímavé postřehy k tématu nebo k chování participantů. Rozhovory se odehrávaly většinou v neformálním prostředí a měly i neformální atmosféru. Tato strategie měla za úkol minimalizovat riziko participantovi nervozity, která by mohla nějakým způsobem ovlivnit jeho odpovědi.

Respondenti

Jak je napsáno výše, výzkumu se zúčastnilo 10 lidí – devět mužů a jedna žena ve věku 19 až 56 let. Důvodem této nevyrovnanosti pohlaví je, že pouze jediná respondentka na základě filtrovacího dotazníku odpovídala předem nastaveným kritériím výzkumu. Vzhledem k zaměření výzkumu toto proporcionální rozdělení nevadí, neboť výsledky šetření nejsou stejně přenositelná na populaci.

Po získání všech rozhovorů jsem respondenty na základě jejich odpovědí rozdělil do tří logických celků, které odpovídají jejich vnímání umělé inteligence (vnímání *strojovosti*, *personifikace* a *nehmatatelnosti*[27]*),* čímž se stala následná práce se získanými daty přehlednější. Aby nebyla četba mé práce tak abstraktní, rozhodl jsem se respondenty anonymizovat přidělením jiných jmen podle abecedy (A–K). Zde jsou respondenti, kteří se zúčastnili mého výzkumu:

[27] Viz podkapitola *Tři kategorie vnímání IVA* v Analytické části.

1. Aneta – 25 let, absolventka VŠE
2. Bedřich – 31 let, konstruktér
3. Cyril – 28 let, datový analytik
4. Daniel – 28 let, státní zaměstnanec
5. Erik – 27 let, auditor
6. Filip – 23 let, programátor
7. Honza – 56 let, podnikatel
8. Ivan – 19 let, student gymnázia
9. Jakub – 34 let, bankéř
10. Karel – 42 let, řidič

Analytická strategie

Analýza výzkumných rozhovorů

Analýza výzkumných rozhovorů byla již komplexní a probíhala v několika krocích, které i popisuje Strauss a Corbin. [Str1999] Jako metodu analýzy jsem zvolil kódování. Tato metoda se přesně hodí pro učení nějakých znaků, nebo nalezení skrytého vztahu.

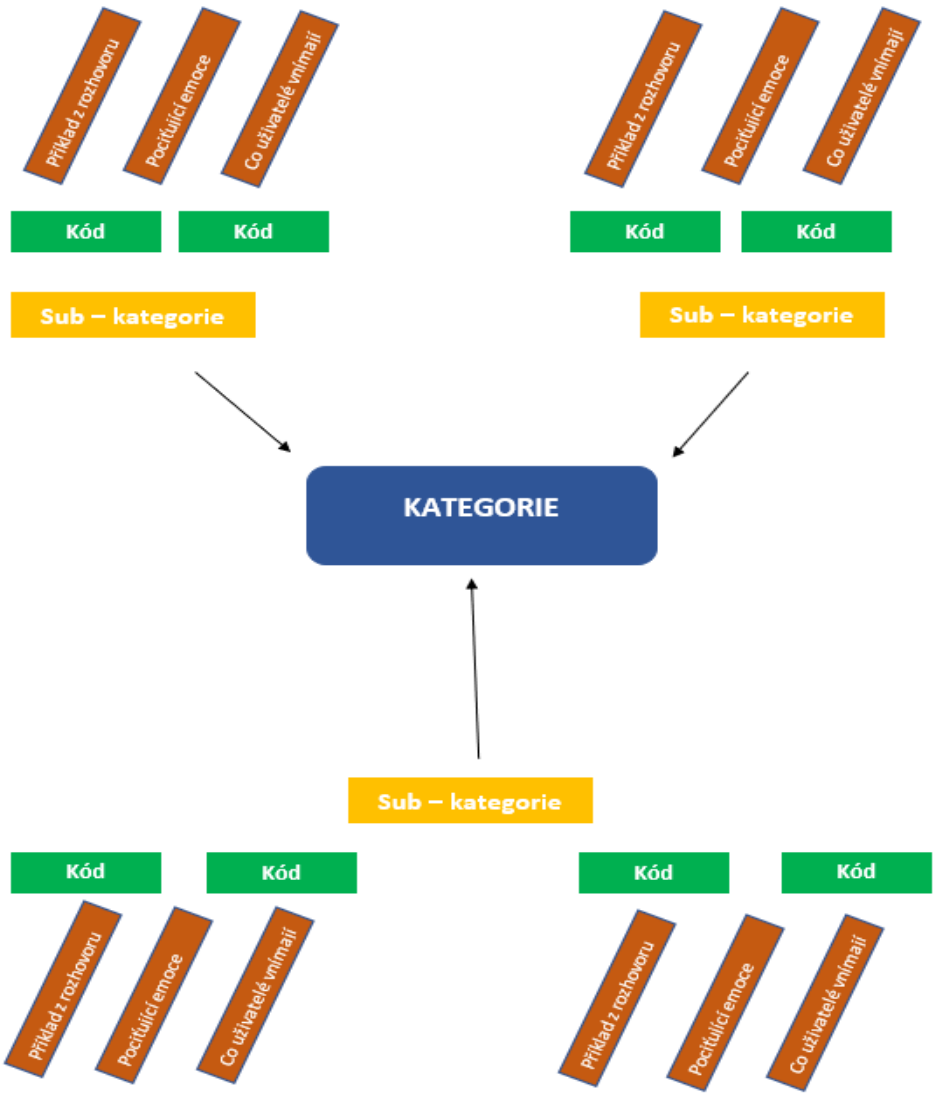
Nejprve jsem provedl otevřené kódování rozhovorů. V přepisech jsem začal postupně hledat důležité informace, či pasáže, které měly nějakou souvislost s daným tématem. Tyto informace jsem označoval a řadil do různých kategorií, které je zastřešují. Také jsem použil metodu kódování in vivo, přičemž informace pojmenovávám tak, jak ji nazval přímo můj participant. Použití kódů in vivo jsem se rozhodl kvůli zachycení emické perspektivy samotného respondenta. Tato analýza otevřeného kódování se poté opakovala do doby, než došlo k empirickému nasycení vzorku, a tedy do doby, kdy výpovědi respondentů již nepřinášeli nové informace. Na potencionální kódy jsem se díval, jak pomocí deduktivní analytické strategie, tak i pomocí indukce. Deduktivní klasifikace výroků participantů má za cíl vyhledávat informace a dávat je do souvislostí s fakty uvedenými v teoretickém rámci, který jsem představil v předchozí kapitole. Na základě tohoto kroku jsem tak získal první pohled na vzájemné vztahy mezi jednotlivými faktory. [Harper, Thompson: 2011] Tím jsem také došel k prvním sociální reprezentacím.

Při indukční analytické strategii jsem postupoval zobecňováním empirických protokolárních vět z výzkumných rozhovorů, které jsou objektivním popisem empirické zkušenosti. Jinými slovy jsem na základě kódů z rozhovorů vytvářel kategorie a sub-kategorie, do kterých jsem jednotlivé kódy

třídil, přičemž jeden kód mohl být i ve více kategoriích. V druhém kroku jsem přistoupil k axiálnímu kódování. Již nalezené kódy jsem přezkoumával a napříč přepisy rozhovorů a hledal jsem opakující se znaky. Jednotlivé kódy, vniklé při předchozím kódování musely být v některých případech zobecněny, či naopak konkretizovány. Díky axiální metodě kódování se tak začaly objevovat různé vztahy a podobnosti mezi jednotlivými přepisy rozhovorů.

Nakonec jsem provedl ještě tematickou analýzu za účelem prozkoumání a ověření nalezených vzorců v datech rozhovorů. Tematická analýza poskytuje systematický způsob pozorování nejenom zjevného, ale i skrytého smyslu v rámci zdrojů dat. Zachycuje aspekty, které lidé přebírají, když uvažují o nějakém riziku a jak strukturují lidské myšlení. [Harper, Thompson 2011: 209–223] Výsledkem bylo sloučení, či redukce duplicitních, či nerozvíjejících kódů a sub-kategorií do finální podoby – viz *schéma 1*, které uvádím jako příklad.

Schéma 1: Příklad analýzy rozhovorů



Zdůvodnění metody zkoumání

Zkoumání umělé inteligence z pohledu sociálních věd probíhá již několik desetiletí. První sociologické práce píší o umělé inteligenci již v 50. letech minulého století. Přesto jejímu zkoumání není věnován dostatečný prostor, což se projevuje nedostatečně rozvinutým teoretickým rámce celého konceptu umělé inteligence. Existuje mnoho definic umělé inteligence, ale sociologové ani sami tvůrci inteligentních systémů se nemohou shodnout, zda současný stav úrovně těchto programů a algoritmů se dá považovat za umělou inteligenci v pravém slova smyslu, či jde o pouhou její předzvěst. Dosavadní sociologická poznání tak mnohdy naráží na kritiku jiných autorů z důvodu mylného pojetí umělé inteligence, která se neslučuje s jejich pojetím. [Rezaev, Tregubova 2018: 93–97] Navíc rychlost inovací a posouvání úrovně moderních technologií se neustále zvyšuje. Pro oblast sociálních věd je obtížné zkoumat určitý aspekt umělé inteligence ve společnosti, když se schopnosti umělé inteligence neustále vyvíjí. [Rezaev, Tregubova 2018: 93–97]

Polo-strukturovaný rozhovor přináší řadu výhod, které jsou pro můj výzkum důležité. V první řadě je to propojení zažitých konstruktů o fungování a chápání umělé inteligence v sociální teorii (viz *Teoretická část*) a vlastních zkušeností participantů. [Galletta 2013: 45] Jedná se pravděpodobně o můj klíčový aspekt při výběru metody zkoumání. Teoretické koncepce důvěry jsem propojil s teorií sociálních reprezentací, která je do jisté míry teorií sociálního konstruktu, neboť mimo jiné popisuje, jak vznikají sociální konstrukty ve společnosti. Konfrontací teorie s reálnými zkušenostmi mých participantů, tak získám potřebné informace k zodpovězení mých výzkumných otázek. Přesně dané otázky a zároveň určitá volnost polo-strukturovaných rozhovorů je ideální metodou pro zjišťování subjektivních interpretací respondentů a zároveň oporou, která mě drží ve vymezeném rámci teorie. [Galletta 2013: 46–52] Polo-strukturované rozhovory jsou navrženy tak, aby výzkumník mohl zjistit subjektivní reakce osob na konkrétní situaci nebo na konkrétní jev, který zažili. [Morse, Field: 1995].

Výběr právě polo-strukturovaného rozhovoru vychází z jeho flexibility. [McIntosh, Morse: 2015] Důvodem volby tohoto výzkumného rozhovoru je, že otázky jsou předem dány a mají otevřený charakter. Díky tomu všichni respondenti odpovídají na stejné otázky. To na jednu stranu významně redukuje vliv tazatele na vývoj rozhovoru a odpovědi participanta, což usnadňuje také samotnou analýzu rozhovorů. Na druhou stranu, jak jsem již zmínil výše, je nutné alespoň částečné aktivní zapojení tazatele, neboť během rozhovoru může dotazovaný odpovědět na několik otázek najednou a přesné držení se scénáře by pak bylo kostrbaté a neproduktivní. [Hendl a Remr, 2017] Nicméně, výsledná data mohu v rámci analýzy srovnávat. [McIntosh, Morse: 2015] Takže by při správném použití metody a správné analýze budou má data relevantní a účelná. Polo-strukturovaný rozhovor také zajišťuje v jeho průběhu dostatečný kontakt s participantem, kterýžto je důležitý pro získání

důvěry respondenta v tazatele. [Bartholomew, Henderson, Marcia 2000: 286–312]

Teoreticky nabízí tento výzkum i jiný pohled na zkoumání lidského rozhodování a lidského uvažování. Na rozdíl od metod strukturálního funkcionalismu, nebo post-strukturalismu využívá sociální reprezentace referenční rámce a znalosti samotné populace k vysvětlení konkrétních jevů ve společnosti. Nabízí tak konkrétně vymezený pohled na úrovni struktur. [Smith, Joffe 2012: 29]

Nevýhody zvolené metody

Kvalitativní metodologický přístup dává zkoumání vnitřních motivací člověka smysl. Stejně tak, jako tomu bylo v případě výzkumu Clodhny O'Connor, která použila teorii sociálních reprezentací k analýze veřejného vnímání irské recese. [O'Connor 2012: 457–461] Přesto však má tato metoda výzkumu i svá úskalí. Nejenom že tazatel nesmí ovlivnit odpovídání a uvažování svého participanta, aby nedošlo ke zkreslení dat. Ale zároveň musí hlídat, aby se z polo-strukturovaného rozhovoru nestal hloubkový rozhovor. Ten dává sice participantovi velký prostor pro odpovídání a dává mu také prostor a možnost o dané problematice uvažovat z různých hledisek. Nicméně je pak takovýto rozhovor velmi časově náročný na analýzu. Výsledem je velké množství „drobných dat", které jsou relevantní pouze v případech zkoumání drobných nuancí. [McIntosh, Morse: 2015] V případě mých výzkumných cílů by však šlo převážně o plýtvání času tazatele i participanta.

Pro svůj výzkum mám k dispozici poměrně malý vzorek participantů. Omezený počet informátorů mi sice poskytuje dostatečné informace o vnímání umělé inteligence, avšak pouze pro úzký okruh lidí. To znamená, že výsledná data nejsou reprezentativní pro celou populaci a platí pouze pro vymezenou skupinu zkoumaných participantů – tedy pro uživatele inteligentních virtuálních asistentů.

Etika výzkumu

Definice etiky spočívá v tom, že etika se zabývá myšlenkou dobra a předcházení možným škodám. [Aluwihare-Samaranayake 2012: 65–66] Problémy v kvalitativním výzkumu bývají jemnější než problémy v kvantitativním výzkumu. V případě kvalitativních metod hrozí problémy v momentě, když výzkumný pracovník vstoupí do určité komunity a nějakým způsobem na ni začne působit. [Orb, Eisenhauer, Wynaden: 2001] V mém případě jsem však nemusel mnoho etických otázek řešit. Mezi participanti nebyl nikdo, kdo by se řadil do některé z eticky zranitelných skupin, jako jsou děti a dospívající, lidé s psychickými, zrakovými, či kognitivními poruchami. Stejně tak se nejedná o žádné příliš kontroverzní téma a odpovědi mých respondentů by neměli vyvolat žádné společenské pohoršení. Přesto bylo třeba zvážit, zda informátora neupozornit, aby zvážil důsledky svého rozhodnutí se průzkumu účastnit, a dále, aby promyslel i veškeré důsledky, které by jeho odpovědi

mohly přinést. [Shamoo, Resnik: 2009] V mém případě se jedná o důsledky, zda by mého participanta nemohl někdo poznat, na základě jeho odpovědí. Já jsem však tuto informaci explicitně svým respondentům nesděloval, a to z důvodu zachování určité „přátelské“ atmosféry, která usnadňuje získávání informací od respondenta, a také jsem žádného respondenta nenutil odpovídat na otázky, na které nechtěl. Etickou stránku mého výzkumu jsem tak vyřešil pomocí informovaného souhlasu[28], ve kterém jsem respondenty informoval o účelu práce, o analýze a následném zveřejnění získaných dat, o způsobu anonymizace jejich jmen, či o jejich právu z výzkumu kdykoliv odstoupit.

Z etického hlediska je také potřeba vzít v úvahu badatele, jako takového. Aluwihare-Samaranayake upozorňuje, že v případě kvalitativní metod zkoumání by si měl výzkumník pokládat otázku, zda není pod přílišným tlakem. Tlak na výzkumníka může být na příklad způsoben nedostatkem času, limitací finančních prostředků, či ze strany zadavatele výzkumu. V případě, že je výzkumník pod tlakem a nepřizná si to, mohou tím být výrazně ovlivněny jeho výstupy. Buď nezaznamená důležité informace, anebo je nesprávně pochopí a následně mylně interpretuje. Příčinou takového nátlaku může být časová tíseň výzkumu, nebo nátlak nadřízených, popřípadě zadavatelů výzkumu. V takovém případě však etický kodex navrhuje zpomalit činnost za účelem zkvalitnění výstupů. [Aluwihare-Samaranayake 2012: 70–72] Časovou tíseň připustit v případě mého výzkumu musím. Avšak si nejsem vědom, že by tato tíseň měla nějaký vliv na proces získávání anebo analyzování dat.

[28] Informovaný souhlas je součástí příloh na str. 91

Analytická část

Analýza rozhovorů

V následující části představím poznatky, které jsem získal z analýz polo-strukturovaných rozhovorů, a které jsou klíčem k zodpovězení výzkumných otázek definovaných v předchozí části. Výsledky šetření jsem rozdělil na jednotlivé části, ve kterých se zabývám, jak uživatelé vnímají svého virtuálního asistenta, jaké sociální reprezentace jsou s tímto vnímáním spojeny, jak se proměňuje důvěra v čase a kde jsou její hranice. V neposlední řadě uvádím rizika, která si uživatel uvědomuje při interakci s virtuálním asistentem.

Tři kategorie vnímání IVA

Při vyslovení pojmu „*virtuální asistent*“ se jejich uživatelům může vybavit mnoho rozličných věcí, vlastností, okolností, nebo abstraktních představ, které si se svým asistentem spojují. Tyto asociace s IVA jsou rozdílné napříč uživateli, neboť jsou založeny na předchozích zkušenostech, získaných informacích, možnostech využívání, anebo jen prosté ochoty asistenta využívat. Když moji respondenti teprve začínali používat virtuální asistenty, neměli s nimi žádnou předchozí osobní zkušenost. Veškeré vědomosti o IVA měli pouze zprostředkovaně. Virtuální asistent byl pro ně něco neznámého. Teorie sociálních reprezentací říká, že člověk pomocí reprezentací dokáže klasifikovat jemu neznámé objekty prostřednictvím objektů známých a tím mu vysvětlovat realitu – činí nefamiliární věci familiárními. Zároveň však také realitu utváří. [Moscovici 1988: 211–250]. Na základě této teorie tak lze sledovat, jak se uživatelovo vnímání tvořilo a proměňovalo v čase.

Vytvořil jsem tři kategorie, které odrážejí různé pohledy uživatelů na svého virtuálního asistenta. Jedná se o tři odlišné typy vnímání asistenta uživatelem: *strojovost, personifikace a nehmatatelnost*, přičemž každý typ má svá specifika, kterými se odlišuje od zbylých dvou kategorií. Neznamená to však, že by uživatel vnímal virtuálního asistenta čistě pouze jedním z pohledů. Téměř vždy je to syntéza všech tří typů vnímání. Na základě sociálních reprezentací však uživatel jednu z kategorií upřednostňuje a ta utváří jeho pohled na IVA. Tyto kategorie by také mohly být do jisté míry odrazem vnímání virtuálních asistentů určitými sociálními skupinami. Jelikož členové nějaké sociální skupiny mývají podobné sociální reprezentace, jejich vnímání reality by tak mohlo být podobné. Například pro programátory a lidi, co se pohybují v IT, by byla typická kategorie *nehmatatelnost*, neboť si pod pojmem virtuální asistent, či umělá inteligence představí spíše programovací kód, než humanoidního robota ze sci-fi. Kategorizace sociálních skupin však není náplní této práce.

Virtuální asistent vnímaný jako stroj

První kategorii, jak respondenti vnímají svého virtuálního asistenta, lze nazvat jako „*Strojovost*“. Uživatelé si pod pojmy virtuální asistent či Siri představují stroj, konkrétní technologii, která mechanicky plní uživatelem zadané úkoly. Přestože jsou IVA prezentováni výrobci a distributory jako inteligentní pomocníci, například Siri [www.apple.com/siri], uživatelé je vnímají stále jako pouhý stroj bez vědomé přítomnosti nějakých emocí směrem k asistentovi.

„Já ho vnímám pouze jako stroj. Jako tady ten telefon, co leží na stole a když mu řeknu Hey Siri, tak na mě začne reagovat.“ (rozhovor Aneta)

Výrok naznačuje, že vnímání virtuálního asistenta uživateli je úzce spojen s formou, jak asistent reálně vypadá. Pro řadu uživatelů je Siri spojena s mobilním telefonem, který také vnímají jako pouhou technologii (stroj). Virtuální asistent se tak pro ně stává integrovanou součástí dané technologie a pro uživatele má tak mechanickou podobu. Při analýze rozhovorů se vynořovaly kódy jako: *Mobilní telefon*, *Součást telefonu*, či *Aplikace*. Všechny tyto kódy jsem sloučil do kódů *Stroj*, jakožto mechanická věc, anebo do kódu *Aplikace*, kdy existuje přímá návaznost na telefonní zařízení. Do kódu *Stroj* jsem zařadil i například Alexu od společnosti Amazon, která je určená pro ovládání chytré domácnosti a její podoba připomíná malou krabičku nebo stolní reproduktor.

„Je to prostě taková krabička, na kterou mluvím a jejím prostřednictvím můžu třeba zhasnout světla, zapnout televizi nebo i rádio v tom mám. Můžu přes to přepínat stanice, anebo támhle můžu ovládat čističku vzduchu.“ (rozhovor Honza)

V případě vnímání IVA jako stroje se v rozhovorech často objevoval kód *Strojovosti*. Strojovostí je myšleno strojové chování asistenta, jehož práce je sice rychlá a efektivní, ale stále mechanicky strojová. Může to být vnímán jako součást domu nebo bytu, či jako součást nábytku „*taková krabička, na kterou mluvím a prostřednictvím ní ovládám byt*.“ Mechanicky tedy provádí příkazy jeho uživatele. Důležité je, že ho uživatel Honza vnímá v kontextu svého bydlení, neživá součást jeho domu. Nevyrovná se lidské dovednosti a nedokáže se vyrovnat s nepředvídatelnými událostmi. Respondenti si pod tím představovali *nedokonalosti* a *neschopnost něčeho*. Dalším, často zmiňovaným projevem strojovosti asistenta jsou *technické nedokonalosti* či *mechanická opotřebení*. Tím je myšleno například postupné zpomalování rychlosti asistenta, špatné pochopení příkazu či nesprávné reakce.

„*Jako, co mě štve, jsou ty její reakce. Já třeba řeknu ‚Hey Siri‘ a ona nic.“* (Respondentka ukazuje pomalé reakce asistenta) *„Hey Siri…, Hey Siri! … Vidíš, teď se chytla.“* (Až na druhý pokus se Siri zapnula. Rozhovor Aneta)

Přesto byly nedokonalosti v řadě rozhovorů omlouvány, a to zejména s odkazem na mechanické opotřebování časem u většiny strojů. Časté bylo srovnávání s opotřebením u mobilních telefonů či

obecně u jakýchkoliv jiných strojů. Podobným způsobem byly omlouvány například i chyby v porozumění příkazu, kdy byl často pronášen argument: *„je to jen stroj.“*

„Když umělá inteligence udělá chybu, tak to pro mě ještě není důvod k nedůvěře. Chyby dělají i lidi, a dokud prostě umělá inteligence dělá míň chyb než lidi, tak jí prostě důvěřuju.“ (rozhovor Filip)

Důležitý prvek, který má vliv na vnímání asistentů jako stroje uživateli, je také zobrazování technologie distributory a výrobci. V České republice je propagace virtuálních asistentů úzce spojena s propagací konkrétních technologií, přičemž se klade větší důraz na samotnou technologii.[29] Provázání technologie s IVA je zřejmé a pro řadu lidí tak může být první asociací pro virtuálního asistenta nějaký konkrétní druh oné technologie – mobilní telefon, chytrá domácnost.

Na druhou stranu kód *Strojovosti* byl často konfrontován uživateli s kódem *Fascinace technologií*. Konfrontace spočívá v tom, že respondenti sice uvádí, že virtuální asistent je pouhý nedokonalý stroj, na druhou stranu právě tato skupina respondentů je nejvíce fascinována možnostmi, které virtuální asistent nabízí. A přestože vnímají IVA jako stroj, v citaci níže je vidět přítomnost určité formy emocí vůči asistentovy.

„ …, že ještě před pár lety, když jsem byl na střední, tak jsme používali tlačítkový mobily, hráli jsme hada a byli jsme šťastní a teď najednou lidi mluví se svým virtuálním asistentem v mobilu a prostě je to fascinující, jak pokročila doba. (…) *Ten pocit, že virtuální asistent může být tvůj mobil, že na tebe mluví mobil, má ženský hlas a je to takový příjemný.“* (rozhovor Bedřich)

Respondenti často poukazovali na schopnosti jejich virtuálního asistenta jako na něco pro ně nečekaného. Vzhledem k jejich silné asociaci se stroji a strojovostí je schopnosti jejich asistenta silně překvapovaly. To má za následek protichůdné pohledy na virtuální asistenty, potažmo na umělou inteligenci obecně. Respondenty totiž fascinuje komunikační úroveň a rozsah asistenta a z rozhovorů vyplývá, že by si respondenti přáli tuto komunikační úroveň ještě rozšířit a více polidštit. Mezi hlavní požadavky patří schopnost empatie, umět naslouchat a poradit, smysl pro humor, či více lidskou podobu.

Tyto požadavky vychází právě z fascinace možností současných technologií, které jsou ale zároveň konfrontovány kulturním odkazem tradice sci-fi[30] v české společnosti. Někteří lidé tak touží po stroji

[29] Například, když se zajímáte o Siri, na internetových stránkách společnost Apple pro Českou republiku naleznete reklamní sdělení směřující spíše na výhody telefonů a tabletů, přičemž možnosti a funkce Siri jsou omezeny spíše na poznámky u daných produktů. www.apple.com/cz/

[30] Tradice sci-fi je původně literární žánr, nazývaný též vědecko-fantastickou literaturou, vycházející nejenom z představ o světě, ale i zkušeností o jeho fungování, inspirovaný pokrokem ve vědě a technice. (…) Vyznačuje se vyjádřením specifické životní filosofie, dramatického vykreslení vize budoucí společnosti, vztahů mezi lidmi a jejich způsobu života. [Maříková, Petrusek, Vodáková a spol. 1996: 969–970].

s lidskými vlastnostmi, ale na druhou stranu se těchto vlastností bojí. Hranice mezi chtěnými a nechtěnými vlastnostmi je mlhavá a neustále se posouvá. Sami respondenti si často během výzkumného rozhovoru uvědomili, že neví, jak k takové technologii přistupovat ani dokonce, jak ji nazývat. Respondent tak na začátku rozhovoru říká, že by umělou inteligenci v uvítal, ale že v současné době to ještě není umělá inteligence, ale pouze lepší stroj.

„…ještě to není umělá inteligence. Ještě si neuvědomují sami sebe." Na druhou stranu po pár otázkách respondent říkal: „*Jako problém to bude, až si dokáží uvědomit sami sebe*. *Protože každý přemýšlející tvor si řekne, že teď sloužím já vám a už mě to nebaví a chci si dělat věci po svém a už bychom se mohli stát nástroji my.“* (rozhovor Bedřich)

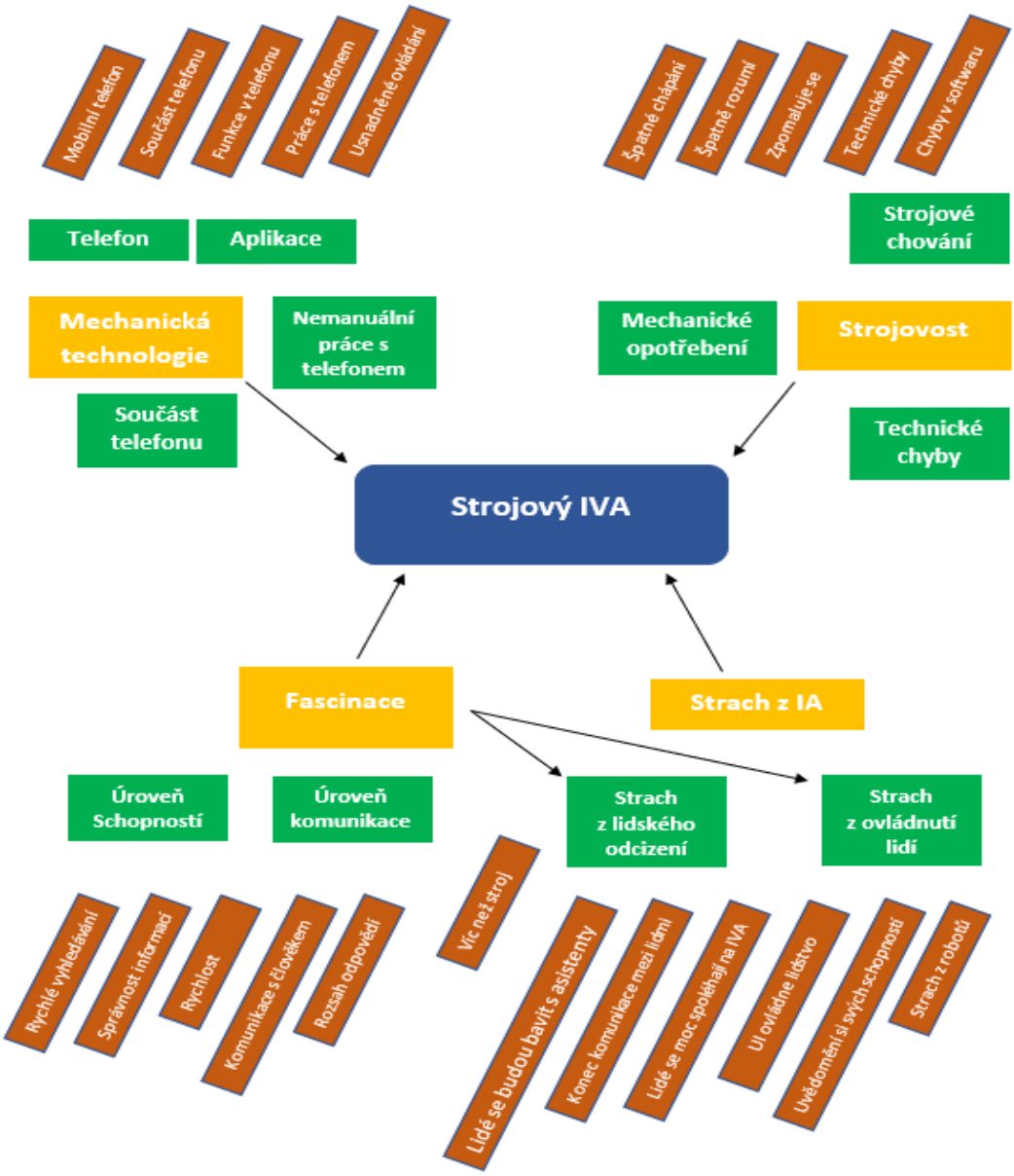
„Chtěla bych, aby mi dokázala pomoc, třeba poradit. Aby byla vtipná, měla smysl pro humor. Mohla by třeba utěšovat lidi, nebo tak.“ Když jsme se však s respondentkou bavili o emocích, její názor se změnil. *„Kdyby se mě třeba najednou zeptala, co mi je, nebo tak, tak to mi přijde děsivý. Že někdo digitální se mnou chce najednou mluvit*. (…) *Jako je to stroj, ale kdyby tam byla ta emoční stránka, tak už by to pro mě nebyl stroj.* (…) *Bylo by to něco zvláštního, něco mezi nebem a zemí.* (…) *Ale rozhodně bych se necítila bezpečně.“* (rozhovor Aneta)

Na těchto citacích je vidět, jak respondenti v průběhu rozhovoru podávají dva různé pohledy na jejich vnímání, které jsou navzájem v přímém rozporu. Když poté byli dotazováni na konkretizaci bezpečných a nebezpečných schopností virtuálního asistenta, respondenti si sami nebyli jisti, kde leží pomyslná hranice mezi tím, co chtějí a čeho se obávají. V případě rozhovoru s respondentem Bedřichem je možné si všimnout, že Bedřich nebezpečí přirovnává k revoluci a nějaké historické zkušenosti. Přestože to nezmínil přímo, z rozhovoru vyplývá, že myslel revoluci jako například povstání otroků ve starověkém Římě, nebo povstání dělnické třídy ve 20. století. *„Už několikrát se stalo, že ti, co sloužili, tak se najednou rozhodli, že už nebudou poslouchat. A že prostě se ta role prohodí.* (…) *No, že vlastně dojde k revoluci.“* (rozhovor Bedřich) Jeho strach pramení z toho, že uvědomí-li si virtuální asistent, že je pouhým nástrojem, mohl by se pokusit o převrácení rolí a udělat nástroje z lidí. Podobný strach je možné pozorovat i ve druhém příkladu rozhovoru s Anetou, kdy přítomnost emocí promění stroj v očích uživatele v něco víc. Stroj s emocemi v tomto případě evokoval „zlé roboty", což respondentka potvrdila tím, že: *„mám strach z vražd*ících robotů, *a že nás mohou někdy ovládnout.“* Tento strach má pak patrně původ v tradici science fiction, protože sama respondentka uvedla, že filmů s podobnou tématikou viděla mnoho.

Kromě strachu z možného vzestupu inteligentní strojů, se často objevoval i opačný strach, a to strach z určité formy lidské degenerace z důvodů proměny sociálních vztahů. To spočívá v myšlence, že lidé budou nové technologie a stroje natolik využívat, že nakonec dojde k proměně sociálních vztahů a tím k velkému oslabení mezilidských vazeb. Oslabení nebo přerušení velkého počtu sociálních

vazeb poté uživatelé mohou vnímat jako potenciální hrozbu pro lidské přežití. Tento strach patrně spočívá v situacích, kdy lidé tráví na svém mobilním zařízení stále více času a mezilidská komunikace se tak přenáší do kybernetického prostoru. Komunikační čas v kybernetickém prostoru je však uživatelem distribuován nerovnoměrně mezi ostatní aktéry sociální sítě. Tato nerovnováha tak způsobuje oslabování jeho širších sociálních vazeb. [Mirilello a spol. 2013: 89–95] Také se často objevoval strach z oslabení lidských dovedností a přílišného využívání asistentů a jiných technologií. Obavy pramenily z pocitu přílišné lidské závislosti na technologiích.

Schéma 2: ***Vnímání IVA jako stroj***



Personifikace virtuálního asistenta

Jak už jsem naznačil dříve, pro různé uživatele, či sociální skupiny má virtuální asistent jinou podobu. Někteří mohou IVA vnímat jako stroj, který plní jejich povely. Mají k němu neutrální vztah a sami si nejsou vědomi nějakých emocí při komunikace s ním. Naopak v této kategorii jsou asistentovi uživatelem připisovány různé lidské vlastnosti a postavení, na jejichž základě dochází k definování vzájemného vztahu uživatele a asistenta. Schopnosti virtuálního asistenta jsou na takové úrovni, že ho uživatelé nevnímají jako pouhý stroj, ale jako něco víc. Připisují mu například lidské vlastnosti, nebo ho srovnávají s lidskými pracovníky.

„Je to takový můj asistent. Taková sekretářka. Když něco potřebuju, prostě jí to řeknu a ona to udělá. Nebo něco potřebuju vědět a ona i to najde a poví mi to. (…) *Je to jako bys měl přidělenou non-stop sekretářku.“* (rozhovor Jakub)

„Je tam vztah šéf – sekretářka, nadřízený – podřízený. Pán a otrok. (…) *Já mu dám příkaz a očekávám, že to splní.“* (Rozhovor Cyril)

(Vztah) *„…je vlastně, dá se říct, že profesionální. Jako kdybych měl profesionální asistentku.“* (rozhovor Daniel)

Na základě odpovědí je možné sledovat, že respondenti dali svým asistentům různé lidské tituly, kterými definovali svůj vzájemný vztah -> kód *Lidská role*, tedy připodobnění k různým rolím, které definují vzájemný vztah uživatele a jeho asistenta. Jejich vzájemný vztah je definován na základě předchozích zkušeností uživatelů s podobně reagujícími subjekty. To znamená, že vzájemná interakce mezi IVA a uživatelem evokuje uživateli dojem, že mluví s lidskou bytostí. Rozdílná úroveň podřízenosti a nadřízenosti také naznačuje rozdílné vnímání vzájemné spolupráce. Respondent Daniel uvedl, že interakce s jeho virtuálním asistentem vypadá jako profesionální pracovní vztah. Je zde tedy kladen důraz na věci spojené s profesionalitou – rychlost, odbornost a kompetentnost. Na druhou stranu respondent Cyril popsal vzájemný vztah tak, že virtuálního asistenta považuje spíše za otroka. Historické konotace slova otrok mohou naznačovat důraz na kontinuální pracovní vytížení subjektu na veškerou rutinní práci bez jakýchkoli námitek a odmlouvání. Navíc přirovnání k otrokovy poukazuje na fakt, že všichni respondenti, ať primárně vnímají IVA jakkoliv – stroj, personifikovaně, či nehmatatelně, považují virtuálního asistenta za něco podřízeného. Nebo jinak řečeno, že lidé jsou a vždycky musí být virtuálnímu asistentovi nadřazení. K potřebě uživatelů mít jasně vymezený hierarchický vztah se svým IVA se v průběhu analýzy rozhovorů ještě několikrát dostanu. V kategorii personifikovaného vnímání virtuálního asistenta je však myšlenka hierarchického vztahu nejzřetelnější.

Na druhou stranu, ne vždy se tato hierarchie projevuje stejně. Přesto, že Cyril považuje svého virtuálního asistenta za otroka, je si vědom vzájemného působení, které ovlivňuje jejich chování. Znamená to, že i vzájemně hierarchický vztah se v čase proměňuje – jak ukazuje následující citace.

„Já si ho vychovávám. (…) *...myslím, že má vychovaného i on mě. Že ten vztah je takový vzájemný.* (…) *...dokážu položit otázku, tak aby ji pochopil. Už je tu, možná, nějaké vzájemné porozumění.“* (rozhovor Cyril

„Musel jsem se učit volit taková slova, aby mi rozuměl. Ne aby mě chápal, aby mi rozuměl. (…) *Chvilku jsme si na sebe zvykali, ale teď už myslím, že dobrý.* (…) *...a samozřejmě malinko kalkuluju s tím, co budu kdy dělat a co můžu nechat na Bixbym.“* (rozhovor Jakub)

Je vidět, že sami uživatelé si vzájemný vztah uvědomují. Uvědomují si proces učení jeden od druhého. Na základě tohoto procesu následně dochází k úpravě uživatelovy komunikace, potažmo i k úpravě jeho chování za účelem lepší a efektivnější spolupráce s IVA. Je však důležité upozornit, že jejich vzájemný vztah je spíše pracovní, než že by byl založený na emocích. Také stojí za povšimnutí, že respondent Jakub říká svému asistentovi jménem. Ostatní respondenti během rozhovoru také používali jména jako Siri nebo Alexa. Často je však zaměňovali i za jiná označení, jako například za telefon, nebo za zájmeno „ono“. Pouze respondent Jakub během celého rozhovoru mluvil o svém asistentovi jako o Bixbym. Kromě toho na mě během celého rozhovoru působil, že v Bixbym vnímá určitou osobní identitu, která ho polidšťuje.

Většina požadavků, které uživatelé svým asistentům dávají, se týká rutinních prací, které nechtějí provádět sami. Asociace se podřízenými, či otroky, je tak nasnadě. Vzhledem k rozsahu úkolů se dá také mluvit i o určité formě opečovávaní. Nicméně, je potřeba dodat, že současná forma pečovatelského vztahu je čistě bez emoční.

„Slyšel jsem, jak virtuální asistent říká, jaký je venku počasí a co si třeba oblíknout na sebe.“ (rozhovor Daniel)

„Líbí se mi, že prostě zhasne světlo a já se nemusím zvedat, nebo že prostě dělá věci za mě.“ (rozhovor Honza)

„Mohla by umět mě třeba utišit, nebo poradit.“ (…) *...simulovat emoce může. Ale reálný emoce bych asi nechtěl.* (…) *Simulovat emoce ano, ale reálné určitě ne.* (rozhovor Cyril)

V poslední citaci opět dochází k mírnému konfliktu očekávání. Na jedné straně si uživatel přeje, aby byl jeho virtuální asistent empatický a na druhé straně však nechce, aby měl reálné emoce. Podobná očekávání, která jsou ve vzájemném rozporu, jsem popisoval i v předchozí kategorii. Podle respondentů může simulování emocí přiblížit IVA blíže k chování lidí a tím jim lépe porozumět. Přirozené emoce jsou však uživateli vnímány jako potenciálně nebezpečné.

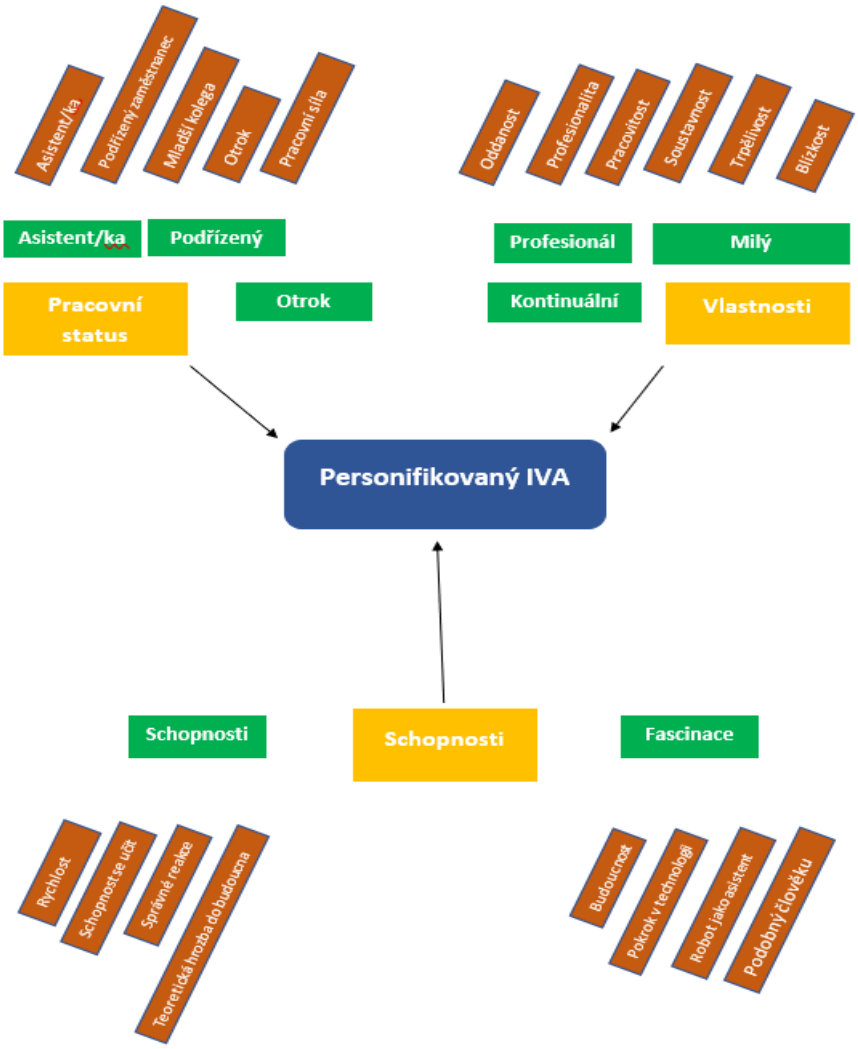
„Myslím, že emoce jsou za mě zbytečné a nežádoucí.“ (rozhovor Daniel)

„…jako určitě bych nechtěl, aby se na mě Bixby začal rozčilovat a třeba mi nadávat.“ (rozhovor Jakub)

Tím se však dostávám i k situacím, kdy je virtuální asistent připodobňován k potenciálně nebezpečnému robotovi. U respondentů, kteří mají tendence svého asistenta více personifikovat, není strach z potenciálních schopností virtuálních asistentů tak velký jako například u lidí, kteří je považují spíše za stroj. Nicméně, v případě, že by IVA mohl mít emoce, respondentům by to nahánělo určité obavy. I v tomto případě pochází strach z kulturní tradice popisování robotů v budoucnosti.

„Určitě jsme bombardovaný z filmů a knih o tom, že prostě se tady ty vynálezy obrátí proti nám.“ (rozhovor Cyril)

Schéma 3: Vnímání IVA jako živou entitu



Nehmatatelný virtuální asistent

Zatímco v předchozích dvou kategoriích měl virtuální asistent pro uživatele vždy hmatatelnou podobu, v případě odpovědí respondentů v této kategorii, je tomu naopak. V tomto případě vnímají respondenti IVA jako něco nehmotného nebo nekonkrétního. Pod pojmem virtuální asistent si představují spíše zdrojový kód, způsob programování, program, algoritmus a další pojmy vztahující se k samotným principům fungování umělé inteligence. Nejlépe to vyjadřuje následující citace z rozhovoru.

„Mně přijde, že je to takové zkreslené, že si všichni představujeme roboty a podobně. Tu představu nemám zhmotněnou. Pro mě je to ten kód a ten program. A přemýšlím nad tím, jak se může chovat.“ (rozhovor Erik)

U odpovědí respondentů, které byly zařazeny do kategorie *Nehmatatelný virtuální asistent*, nejvíce dominovaly kódy – a*lgoritmus*, *kód*, *program* a *programování*. Tedy s pojmy, které souvisí s principem fungování virtuálních asistentů, anebo souvisí s procesem jejich vytváření. Respondenti se tak většinou dívali na fungování IVA potažmo umělé inteligence z pohledu lidí, jenž takové technologie vytvářejí. Vzhledem k jejich abstraktnějšímu vnímání virtuálních asistentů než u předchozích dvou kategorií, je také spektrum jejich odpovědí širší. Jedním z projevů jejich rozsáhlejšího vnímání IVA je, že během rozhovorů uváděli mnoho oblastí, kde by mohli být virtuální asistenti bez větších obav implementováni, či jak dalece by mohli být využíváni.

„Já si vždycky vzpomenu na počítačové hry. Jsem hráč počítačových her a tam ta umělá inteligence hraje docela velkou roli.“ (rozhovor Filip)

„Myslím, že asistenti by mohli být kdekoliv. Mně se velmi líbí koncept chytrého města. Viděl jsem pár návrhů, možností… Myslím, že tam všude by našli asistenti a umělá inteligence obecně široké uplatnění.“ (rozhovor Erik)

Během rozhovorů se pravidelně objevovaly dva střídající se motivy vnímání technologie. První motivem byla častá kritika úrovně inteligence a nedostačujících schopností současných virtuálních asistentů. Kritika se také týkala nedostatečné implementace do každodenního života a míry jejího využívání. Zdroj kritiky současného vztahu spočívá ve znalostech programování a osobních zkušenostech s IT technologiemi. Jako zdroj informací často uváděli: *odborná školení*, *odborné články a videa*, *pracovní zkušenosti*, či *zájem o nové technologie*. Jejich zájem o nové technologie se u respondentů projevoval technooptimismem a vidinou fascinující budoucnosti, který v nich posiloval touhu po nových technologiích. Tento svůj postoj dávali během rozhovorů do přímého kontrastu s kritikou současné úrovně technologie. U respondentů se objevovala vidina fascinující[31]

[31] Sami respondenti často používali slovo *„fascinující“*.

budoucnosti a moderních technologiích, která v současném světě ještě není možná, ale vlivem nových objevů a technologií, jednou možná bude.

„Já jí právě důvěřuju, protože si myslím, že právě není dost dobrá, a ještě hodně dlouhou dobu nebude dost dobrá. (…) *V podstatě každá práce by mohla být nahrazena umělou inteligencí, kdyby ta umělá inteligence byla dost dobrá.* (…) *Umělá inteligence může za člověka udělat něco, co člověk, buď ani nedokáže v některých případech udělat, anebo to udělá prostě rychleji.“* (rozhovor Filip)

„On je to vlastně v uvozovkách hloupý přístroj, který se učí, ať už od tebe, nebo ho někdo učí, jak má fungovat. (…) *Ono to vlastně souvisí s rozvojem 5G sítě, kdy vlastně teď není ta kapacita na to tu technologii dál rozvíjet, nebo je ta kapacita malá. Ale potom, až ta 5G síť bude všude, tak vlastně ty možnosti se zněkolikanásobí.* (…) *Líbí se mi propojení technologie a každodenního života.* (…) *Líbí se mi koncept chytrého města.“* (rozhovor Erik)

„Já si nemyslím, že asistent je úplně umělá inteligence. Sice je to tak všude prezentováno, ale otázkou je, jestli umělá inteligence vůbec existuje. Pořád je to jen naprogramovaný algoritmus, kterému řeknu, jak se chovat. (…) *Všechno to závisí na výkonu výpočetní techniky…. Na výkonu počítače, když to řeknu zjednodušeně.* (…) *Dovedu si představit, že virtuální asistenti budou na každém rohu, místo prodavačů* (...) *budou všude, ale budou to fakt spíš jen asistenti.* (rozhovor Ivan)

Na výše uvedených rozhovorech je vidět, jak respondenti uvažují nad virtuálními asistenty z pohledu tvorby jejich schopností – programování, kódy, algoritmy. Uživatelé sledují možnosti IVA v kontextu vyspělosti související technologie a právního rámce. Dle některých respondentů není právní rámec na implementaci virtuálních asistentů do každodenního života v masovém měřítku vůbec připravený.

„ …samozřejmě otázka, zda jsme na to připravení. Já si myslím, že ne.“ (rozhovor Filip) *„Jo, ale ten algoritmus nikdy není dokonalý. A když se někomu něco stane, kdo bude vinen?“* (rozhovor Cyril) *„Připraveni na to asi jsme, ale určitě ne právně.“* (rozhovor Erik) *„Muselo by se to silně právně ošetřit.“* (Rozhovor Ivan)

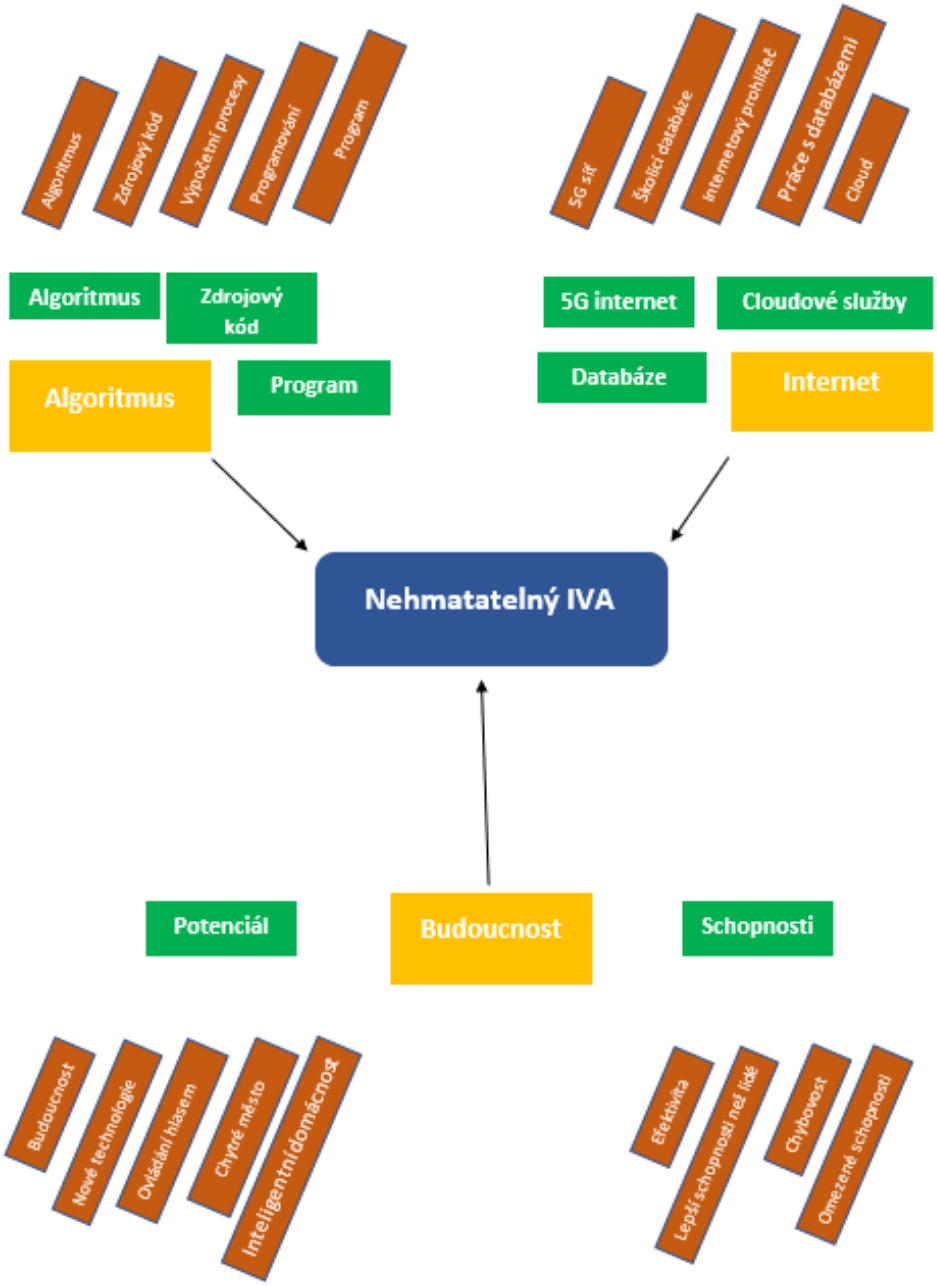
Znalosti respondentů o umělé inteligenci vychází z jejich zájmu o možnostech a schopnostech fungování současných technologií. Zajímají se o a principy UI a prostřednictvím těchto znalostí vnímají nově vzniklé technologie. Jak jsem psal výše, jejich znalosti čerpají primárně z odborných zdrojů, článků na internetu, či vědecko-populárních časopisů. Často mívají i praktické zkušenosti s nějakou formou programování. Při analýze jsem použil kódy: *odborná školení*, *odborné články a videa*, *pracovní zkušenosti*. U této kategorie respondentů, na rozdíl od předchozích dvou kategorií, se téměř neobjevoval strach z vysokých schopností virtuálních asistentů, potažmo umělé inteligence. Naopak se zde často objevovali odpovědi, které si nepřipouštěly možnost nadvlády umělé

inteligence. Opět názory respondentů vycházely z jejich předchozích zkušeností a jejich znalostí z odborně populárních zdrojů.

„Ale nepatřím k lidem, který by měli strach z toho, že umělá inteligence někdy ovládne lidstvo.“ (rozhovor Filip)

„Obecně mi přijde, že všichni tihle autoři tady toho sci-fi, mi přijde, že jsou strašně optimistický v tom, že si myslí, že budeme mít za 20, 30 let inteligentní umělou inteligenci. Já si myslím, že nikdy nebudeme mít inteligentní umělou inteligenci.“ (rozhovor Erik)

Schéma 4: Vnímání IVA jako něco nehmatatelného



Formování důvěry v IVA

Pro účely zjištění zdrojů důvěry k virtuálním asistentům jsem se rozhodl do svého výzkumu vybrat pouze ty respondenty, kteří ve filtrovacím dotazníku uvedli, že umělé inteligenci důvěřují nebo spíše důvěřují. Z předchozí části však vyplývá, že pro každého respondenta může mít jeho virtuální asistent jinou podobu. Někdo IVA vidí jako stroj, někdo mu dává personifikované vlastnosti a někdo za ním vidí programovací kód. Z toho vyplývá, že přestože všichni respondenti důvěřují umělé inteligenci, potažmo inteligentním asistentům, zdroje jejich důvěry jsou odlišné. V této části rozeberu a analyzuji ty části rozhovorů, kde respondenti hovoří o důvěře ve virtuální asistenty, o počátcích jejich důvěry a o limitech důvěry.

Počátek a prohlubování důvěry

Tabulky 2: Počátek důvěry u respondentů

Respondent	Počáteční důvod využívání
Aneta	*„Dostala jsem inteligentní hodinky, a tak jsem začala využívat Siri, abych nemusela pořád hledat v kabelce telefon a mohla všechno ovládat přes ty hodinky.“*
Bedřich	*„Byl jsem zvědavý, jak to funguje.* (…) *Jestli mi to usnadní život.“*
Cyril	*„Viděl jsem nějaký video, že to jde použít, nebo nějakou zprávu, že to takhle funguje, tak jsem to taky zkusil.“*
Daniel	*„Viděl jsem to prostě na nějakých videích, přišlo mi to super, tak jsem to taky zkusil. No a přišlo mi to zajímavý.“*
Erik	*„Zájem a neznalost nové technologie. Zkusit si to, jak to v praxi funguje.“*
Filip	*„Reklamy na internetu, že jsem viděl a přišlo mi to cool.“*
Honza	*„Vlastně mě baví nové technologie, a když jsem se dozvěděl, že je možnost vlastně domácnost ovládat přes asistenta, tak jsem to chtěl vyzkoušet.“*
Ivan	*„Chtěl jsem to zkusit.* (…) *Už nevím, kde jsem to viděl, ale zaujalo mě to.“*
Jakub	*„Viděl jsem, že ho používá můj kamarád a hrozně se mi to líbilo.* (…) *Když jsem zjistil, že mám Bixbyho v telefonu, tak jsem ho zkusil a od té doby ho používám.“*
Karel	*„Někde jsem viděl, že mi to může ulehčit práci s telefonem a vyhledávání, tak jsem to zkusil a ono to fungovalo.“*

Počátečním impulzem pro začátek využívání virtuálního asistenta byla prakticky u všech respondentů zvědavost a fascinace. Většina respondentů v rozhovorech uvedla, že nejprve vnímali svého virtuálního asistenta jako funkci v telefonu, popřípadě jako součást telefonu. Někteří ještě doplnili, že při počátcích jejich interakce s IVA měli pocit, jako že jsou v budoucnosti. Tím mysleli

neurčitou bodu v budoucnu, která je plná nových technologií a nových možností. V Tabulce 2 jsou vypsány odpovědi všech respondentů na otázky spojené s počátky a důvody využívání jejich virtuálních asistentů. Z tabulky je patrno, že důvodem byla u většiny respondentů zvědavost, jak tato „funkce v telefonu“ vlastně funguje.

Vnímání respondentů se však postupem času proměňovalo. Někteří respondenti vnímali svého IVA jako stroj a postupem času mu začali dávat personifikované vlastnosti. V průběhu využívání asistenta tak muselo dojít k situacím, které ovlivnily uživatelovo vnímání asistenta tak, že začal IVA vnímat trochu jinak. Jako jeden z nejdůležitějších milníků v průběhu využívání asistenta, který silně ovlivnil vnímání uživatelů, uváděli respondenti vůbec jejich první interakci s asistentem.

„Ten první kontakt byl takový asi nejvýraznější pocit. (…) *„Ten pocit, že virtuální asistent může být tvůj mobil, že na tebe mluví mobil, má ženský hlas…“* (rozhovor Bedřich)

„*Vůbec, asi ten začátek, že si to člověk zkouší*. (…) *Byl jsem překvapený, co to všechno umí.“* (rozhovor Erik)

„Asi když jsem s ním mluvil poprvé. (…) *Bylo to něco nového a dělalo to na mě dojem, že jsem si do té doby nedokázal představit, že když mu řeknu nějakou věc, tak mi to je schopné odpovědět.“* (rozhovor Filip)

Na těchto citacích je vidět, že první interakce řadu uživatelů překvapila. Zejména kvůli schopnostem vyhledávání asistenta a jeho úrovně komunikace. Respondenti na tyto dvě schopnosti často upozorňovali. Interakce s asistentem byla co do očekávání natolik uspokojivá, že se respondenti rozhodli ji i nadále využívat. Zkušenosti v průběhu využívání IVA tak mají důležitý vliv na důvěru a na její proměnu v čase. Následující citace, které budu popisovat, jsou odpovědi na otázky: Kdy Vás váš virtuální asistent překvapil? (Pozitivně/Negativně) Jaký je nejsilnější zážitek s Vaším virtuálním asistentem a podobně. Cílem těchto otázek je najít zlomové okamžiky v průběhu interakce, které nějakým způsobem ovlivnili vztah uživatele k asistentovi, a tím i důvěru k němu.

„Že jsem se ho zeptal třeba na to počasí a on mi řekl, že je šance, že bude pršet, nebo vypadá to, že nebude pršet vůbec, tak jsem byl nadšený. Že je to super, že mi to takhle přímo řekne. Když se ho zeptám na hodinu, jestli bude pršet, on mi řekne, že bude pršet během dne. (…) *Překvapilo mě, že to všechno umí.“* (rozhovor Cyril)

Uživatel zde popisuje situaci, kdy se svého virtuálního asistenta zeptal: Jaké bude dnes počasí? Jelikož se mu dostalo jasné, srozumitelné, a především přesné odpovědi, rozhodl se respondent Cyril funkci využívat častěji i s možností podrobnějších informací (V kolik bude pršet? / Jak dlouho bude pršet?) Dá se tedy říci, že se v této oblasti prohloubila uživatelova důvěra ke svému asistentovi, jinak by funkci patrně nadále nevyužíval. Navíc ještě uvedl, že informace byla, pro něj, podána takovým

způsobem, o kterém vůbec netušil, že by tohoto byl stroj schopen. Později v rozhovoru byl respondent Cyril dotázán, co rozhoduje o tom, pro jaké činnosti IVA využívá. Cyrilova odpověď odhalila princip, který je pro prohlubování důvěry v průběhu času klíčový, a který používají i další uživatelé.

„Já mu dám vždycky úkol a očekávám, že ho splní. A on ho splní. Tak mu dám další úkol. A takhle pokračuju. (…) *Když něco zvládne a zvládne to víckrát tak mám prostě důvod mu věřit a můžu mu to následně svěřovat bez strachu.“* (rozhovor Cyril)

Na základě osobní zkušenosti, kdy IVA naplnil, nebo dokonce předčil, očekávání svého uživatele, došlo u uživatele k prohloubení důvěry ve virtuálního asistenta a vznikla motivace ho pro tuto úlohu využívat. Při opakovaném naplnění očekávání má pak uživatel tendenci svěřovat svému IVA stále nové úkoly. Uživatel sleduje, jak se virtuální asistent vypořádá s jiným, pro oba novým, úkolem. Uživatel si je vědom možného rizika, na druhou stranu jeho virtuální asistent ho v předchozích situacích nezklamal, a tak má jeho důvěru. Na základě podobných interakcí se tvoří důvěra v asistenta. Na druhou stranu, pokud asistent požadovaný úkol nesplní, nedojde k naplnění očekávání a tím dochází k oslabení důvěry ve schopnosti asistenta, nebo v asistenta obecně. Princip oslabování důvěry funguje principiálně stejně, jako budování důvěry v něj. Když nedojde k naplnění očekávání, uživatel nemá motivaci svěřovat IVA nové úkoly a důvěra se tak nemá jak rozvíjet dál a dochází ke stagnaci, či poklesu.

„Na začátku jsem ho využíval jen ta takové ty klasické věci. Kolik je hodin, jak je venku a podobně. (…) *Vím, že jednou jsem potřeboval napsat e-mail v angličtině, a tak jsem to zkusil napsat přes Bixbyho, aby mi zkontroloval a opravoval gramatiku. No a… Jako, kde jsou dvě „t“ a podobně, takhle. No a on to zvládnul. A ještě mnohem líp než jsem čekal.* (rozhovor Jakub)

I zde je respondentova důvěra rozvíjena na základě naplnění nějakého očekávání, v tomto případě vyzkoušení funkce psaní e-mailu na základě toho, co uživatel diktoval. Respondent Jakub byl s výsledkem natolik spokojen, že se rozhodl tuto funkci i nadále využívat. Najednou si svého IVA začal spojovat s reálným asistentem, kterému diktuje, co má psát. Došlo tak k formování nových sociálních reprezentací, které prohloubili důvěru uživatele ve svého virtuálního asistenta. Toto připodobnění se projevuje pasáží, kde uživatel chce využít virtuálního asistenta na kontrolu gramatických chyb v jeho e-mailu.[32] Na základě vzájemných interakcí tak postupem času tak u uživatelů dochází k proměňování důvěry.

„Vlastně nejdřív jsem s ní ovládal jenom televizi. To byla ještě nějaká jejich aplikace. (…) *… a jelikož to fungovalo, a tak když jsme rekonstruovali ten dům, tak už jsem uvažoval, že si ji koupím.*

[32] Podrobně píši o principech ukládání nových sociálních reprezentací u uživatelů IVA v části *Sociální reprezentace jako podmínka důvěry*.

(…) *Všechno to fungovalo, všechno bylo jednoduší, pohodlnější.* (…) *Jako ono to je pořád na stejném principu, takže když funguje jedno, tak prostě bude fungovat i to druhé. Tady už se spíš bavíme o nějakém problému, jako vadné čidlo nebo rozbité…"* (rozhovor Honza)

Respondent Honza pomocí asistenta Alexy ovládá svoji domácnost. Popisoval, jak využívá možnost připojení některých přístrojů v domě k Alexu (světla, klimatizace, televize, čistič vzduchu). Po dobré zkušenosti s fungováním Alexy se rozhodl ji ještě více využívat. Jedním z jeho argumentů je, že Alexa pracuje stále na stejných principech, a tak ji může využívat bez obav kdekoliv. Alexu bere jakou „součást nábytku", která pracuje kontinuálně pořád na stejném principu a je neoddělitelnou součástí domácnosti. Sociální reprezentace uživatele jsou tak spojený s místy, kde na něj Alexa působí – prostřednictvím světel, televize, klimatizace. Uživatel má tak v tomto případě virtuálního asistenta spojeného především s nábytkem a zároveň s malou krabičkou. Sociální reprezentace spojené s IVA jsou nábytek, krabička, stroj, něco mechanického.

„Tak určitě, jsem třeba i zkoušel v rámci vyhledávání webových stránek. Bylo to zajímavý, nebo mě překvapilo, že to fakt funguje. Že i v češtině, v rámci toho odezírání to nějakým způsobem běží. Samozřejmě že ta angličtina je nejvíce preferovaná, ale že to funguje." (rozhovor Erik)

Respondent zde mluví o vyhledávání na internetu, přičemž byl překvapen, do jaké míry je jeho asistent schopen reagovat na příkazy v češtině a vykonávat je. Z tohoto důvodu později v rozhovoru uvádí, že právě v tom, jak asistenti dokáží už dnes reagovat na povely hlasem, vidí velkou budoucnost, kdy: *„lidé budou všechno ovládat hlavou a ne rukama."* (rozhovor Erik) Jeho sociální reprezentace jsou spojeny s budoucností. Lépe řečeno možnostmi, které budoucnost přináší. Respondent Erik si pod budoucností vybavuje možnost ovládání technologie hlasem. Zde je opět pravděpodobný vliv tradice science fiction, kde často dochází k tomu, že lidé komunikují s přístroji a ovládají je hlasem. Zároveň je to i trend výrobců a distributorů virtuálních asistentů k usnadnění práce uživatelů.

„Já to vlastně využívám proto, abych nemusel sahat na telefon. (…) *Takže já ji vlastně vždycky něco řeknu a ona to udělá.* (…) *…jako třeba, najdi tohle, zapni tenhle song, přeskoč tenhle song, jaký bude zítra počasí, co si mám vzít na sebe, kdy něco nastane a tak.* (…) *Jednou jsem byl ve sprše a chtěl jsem po ní, aby přepnula na další písničku. No a ona to neudělala. Tak jsem to zkusil znovu a zase nic. Tak znovu, znovu. Tak jsem na ní začal křičet, ale ona mě úplně ignorovala.* (respondent se směje) (…) *Takže to mě docela naštvalo, no."* (rozhovor Daniel)

Z této citace vyplývá, že respondent využívá IVA především, jako náhradu manuálního ovládání telefonu. Zjištění, že může telefon ovládat na podobné úrovni, aniž by se ho dotýkat, prohloubilo uživatelovu důvěru v asistenta, neboť je to vlastně virtuální asistent, který telefon ovládá. Mezi uživatelem a asistentem tak vniká určitý vztah. Telefon je pro řadu lidí velmi citlivá a soukromá věc.

Nechat asistenta, aby ho na základě příkazů uživatele ovládal, určitě důvěru buduje. Blízkost však nespočívá pouze na úrovni vzájemného vztahu. Blízkost je myšlená i ve vztahu k vzdálenosti, kdy virtuálního asistenta má neustále na blízku (mobilní telefon má neustále po ruce). V případě, že telefon po ruce není, blízkost mu zprostředkovává právě virtuální asistent, pomocí kterého může uživatel svůj mobilní telefon ovládat. Na druhou stranu, respondent také uvádí negativní zážitek, kdy AVI nereagoval na jeho příkaz a neprovedl požadovanou akci. V tomto případě došlo k nenaplnění očekávání, které mohlo mít za následek pokles míry důvěry. Uživatel má tendenci se na svého virtuálního asistenta spoléhat. Podobné prožitky však míru důvěry negativně ovlivňují. Respondent Daniel také uvedl, že svého IVA vnímá jako asistenta a jejich vztah popsal jako profesionální, tedy že se na něj uživatel může po pracovní stránce spolehnout, ale žádnou jinou komunikaci (činnost) už od něj nevyžaduje a neočekává. Sociální reprezentace uživatele jsou tvořeny především prožitky, kdy asistent plně nahradil při ovládání telefonu uživatele. Uživatel, v tomto případě tedy Daniel vnímá svého asistenta, jako určitého zástupce (čeho/koho?) – což by i vysvětlovalo popis jejich vztahu jako „profesionální“. Důvěra spočívá na očekávání práce, vzájemné spolupráce a vstřícnosti. [Frederiksen 2012: 733–750]

Hranice důvěry

Sociální reprezentace a s nimi spojené zdroje důvěry se nejlépe ukazují na „hranicích důvěry“. Tedy na okraji toho, co je uživatel ochoten AVI svěřit, co už ne. Také zjistit, a jaké k tomu má důvody. Při analýze jsem tuto kategorii nazval Limity důvěry a zkoumal jsem, jak uživatelé přemýšlí o schopnostech IVA a potencionálním nebezpečí spojeným s jeho využíváním.

„Svěřil bych mu nekritické úlohy. To znamená, že bych mu úplně nesvěřil řízení auta. Nebo kontrolu nad mým bankovním účtem. Je tam ta možnost zneužití a také si myslím, že ta technologie ještě není na tolik vyvinutá, aby udělala všechno bez chyby. (…) *Už jsem slyšel o autech se samořízením, které havarovali.* (…) *ten kód, ten algoritmus v tom nemůže být nikdy bezchybný.“* (rozhovor Cyril)

V Cyrilově případě se ukazují hlavní limity důvěry. První se týká zneužití dat. Strach ze zneužití osobních údajů se objevoval v mnoha rozhovorech. Většina lidí však nedokázala určit, co by to pro ně znamenalo, tedy zda by došlo ke ztrátě důvěry, nebo jenom oslabení důvěry. Jinými slovy, zda by i nadále důvěřovali IVA, nebo by to byl dostatečný důvod k ukončení využívání virtuálních asistentů. Respondenti si nebyli totiž jistí, jakou roli by v případném zneužití IVA hrál. Do jaké míry by byl za zneužití osobních údajů zodpovědný. Respondenti si ani nebyli jisti, zda by se případná nedůvěra týkala pouze toho jednoho konkrétního asistenta, konkrétní značky asistentů, nebo IVA celkově. Jelikož se s podobnou situací nesetkali, bližší informace mi poskytnout nemohli. Okódoval jsem podobné odpovědi kódem *zneužití třetími stranami*. Uživatel si spojil s virtuálním asistentem nebezpečí zneužití. Asociace je spojená s počítači, notebooky, tablety, telefony a různými

databázemi, které jsou častým cílem hackerských útoků. Jelikož je virtuální asistent integrován ve výpočetní technice, sociální reprezentace varují uživatele před možností hackerského útoku. Ve druhém případě se respondent bojí chyby v algoritmu a nastavení samotného asistenta. Argumentuje tím, že nelze naprogramovat bezchybný kód. Virtuálního asistenta tak přirovnal k autonomnímu vozidlu, které, ač se prezentuje jako bezpečnější než vozidlo řízené řidičem, již také havarovalo a zabilo člověka. [Preliminary report – Highway HWY18MH010: 2018]

„No, vzpomínám si, že jsem ji ráno zapnula, nebo jsem po ní něco chtěla a ona se mě najednou zeptala: How are you today? To už bylo takový divný, že se mě vlastně můj mobil ptá, jak se mám. To už bylo takový nepříjemný. (…) *No, protože mám strach z těch robotů, a že nás jednou můžou ovládnout… a tak.“* (rozhovor Aneta)

Tato citace popisuje nejsilnější zážitek respondentky Anety, které se jednou Siri zeptala, jak se má. Zde se projevuje vnímání virtuálního asistenta, které je v rozporu s citací v předchozí části. Aneta v předchozí části uvedla, že IVA vnímá jako pouhý stroj bez emocí. Objektivizovala si Siri jako něco, co je uvnitř telefonu (kód *Součást telefonu*). Sociální reprezentace vycházela z podoby, v jaké Siri vidí. Nyní však byla sociální reprezentace konfrontována jinou sociální reprezentací, která vnímá mluvící stroje jako potenciálně nebezpečné roboty. Zdroj této sociální reprezentace pochází pravděpodobně opět z tradice námětů sci-fi.

„Asi bych ji nedával někam, kde může způsobit nějaké velké škody. Jako kde na ní není žádný dohled. (…) *Nechal bych ji asi v té pozici toho informačního asistenta…* (…) *Ale určitě bych ji nesvěřoval žádnou důležitou pozici, kde by měla rozhodovat.* (…) *Mohla by být někde na recepci, kde by radila lidem.“* (rozhovor Ivan)

Zde je vidět, že IVA má pro respondenta především podobu informačního asistenta. Po stránce informací a asistencí mu respondent důvěřuje. Na druhou stranu by mu již nesvěřil žádné důležitější úkoly než pouze ty, u kterých se přesvědčil, že na ně IVA stačí. Respondent vnímá svého asistenta čistě jako pomocnou sílu bez schopnosti udělat nějaké důležité rozhodnutí. Vychází tak ze svých zkušeností, kdy IVA připodobňuje k informacím a poskytovateli těchto informací – například vyhledávač na internetu. Rozhodovací schopnosti však nejsou podle respondenta tak dobré, neboť po technické stránce nikdy nemohou obsáhnout situaci v celém kontextu: *„…jako myslím si, že to nedokáže všechno obsáhnout.* (…) *Těch vjemů, co si musí hlídat je tolik…“* (rozhovor Ivan)

„Ten asistent se může implementovat asi všude. Nenapadá mě oblast, ve který bych ji nechtěl. Má to velký potenciál pomáhat. (…) *Třeba některé věci v autě bych ho nechal dělat* (…) *třeba odemykání a zamykání auta, řazení, různá nastavení v autě, zapnutí motoru taky…, ale nenechal bych ho třeba se rozjet. Třeba tempomat bych ho nechal, ale rozjet bych ho nenechal.* (…) *To už je moc aktivní.“* (rozhovor Daniel)

U respondenta Daniela je jasně patrné, že jeho předchozí osobní zkušenosti a s tím spojené sociální reprezentace mají vliv na limity jeho důvěry k virtuálnímu asistentovi. Výše jsem totiž uváděl jeho citaci, kde píše, že virtuální asistent je pro něj náhrada za manuální práci s mobilním telefonem. Vnímá ho jako asistenta, který na jeho pokyn provádí různé operace na mobilním zařízení (vyhledávání na internetu, pouštění písniček, drobná nastavení v telefonu). Dle jeho odpovědi je vidět, že hranice důvěry ve virtuální asistenty kopíruje jeho činnost se Siri. Respondent Daniel by dal virtuálnímu asistentovi v autě možnost ovládat řadu věcí – řazení, odemykání auta, zapínání motoru. Podobné činnosti nechává nyní dělat svého asistenta v mobilu. Jeho přístup tak vychází ze sociální reprezentace, kde AVI považuje za svého profesionálního[33] asistenta a svěřuje mu poměrně širokou škálu pravomocí. Omezení pravomocí se týká především aktivních zásahů do chodu zařízení – automobilu, nebo telefonu.

Proces ověřování informací a jeho vliv na důvěru

V předchozí části jsem popsal, že důvěra k virtuálnímu asistentovi se může proměňovat. Zásadní roli v procesu hrají sociální reprezentace, které se významně podílí na změně vnímání IVA uživatelem, čímž dochází k jeho redefinování a k posunu důvěry. Součástí tohoto procesu je i ověřování informací, které virtuální asistent uživateli předkládá. V následujíc části ukáži, jak ověřování informací ovlivňuje počáteční důvěru i jak ji postupně může měnit. Také poukáži na možnost vztahu, mezi mírou důvěry a rozsahem využívání virtuálního asistenta.

Procesem ověřování informací mám na mysli postupy, jak se uživatelé přesvědčovali, zda jejich virtuální asistent splnil jejich příkazy. Během rozhovorů jsem se tak ptal, zda si respondenti ověřovali informace a proč. Dále jsem se tázal, jakým způsobem docházelo k ověřování informací, a jak dlouho popřípadě, jak často k ověřování docházelo. Z rozhovorů vyplynulo, že postup u většiny respondentů byl velmi podobný, a tak zde ukáži pouze krátké úryvky popisující přístup a proměnu procesu ověřování informací v čase.

„Na začátku jsem si informace ověřoval… (…) *No, protože je to jen stroj a ten se může mýlit.“* (rozhovor Bedřich)

„Určitě jsem si ověřoval informace. (…) *Jako, neznal jsem to, nevěděl jsem, jak to bude reagovat, jak to bude poslouchat. Věděl jsem, že to jen nějaký program, který nějak funguje, ale jak se bude doopravdy chovat, jsem netušil.* (rozhovor Filip)

[33] *Profesionální* – respondent použil přímo v rozhovoru, viz výše: oddíl *Personifikace virtuálního asistenta*, tedy, že se na asistenta může spolehnout po pracovní stránce (v zadaných úkolech), ale žádné jiné činnosti ani vazby neočekává.

„No určitě. Chtěl jsem vědět, jak to funguje, co to všechno umí, a tak jsem mu dával všemožný úkoly hned od začátku.“ (rozhovor Erik)

„Já jsem mu dal na začátku nějaký úkol a čekal jsem, jestli ho splní. No, a když ho splnil, tak jsem mu dal další úkol.“ (rozhovor Cyril)

„Tak pořád je to technika, takže stát se může všechno. Vůbec jsem nevěděl, jak to bude reagovat, jestli to bude navzájem komunikovat… (…) *Ale zkoušel jsem to dál a ono to šlo.* (rozhovor Honza)

„Zajímalo mě, jaký bude. Dával jsem mu úkoly a hledal ty hranice, co dokáže a co už je moc. (…) *Jo, tak kontrola tam byla, ale byla spíš z toho důvodu, že jsem zkoušel, co všechno dovede.“* (rozhovor Jakub)

Jak je z úryvků rozhovorů patrno, většina participantů nějakou formu kontroly prováděla. Hlavním důvodem byla primárně počáteční neznalost technologie. Uživatelé měli o IVA určité představy, ale žádnou osobní zkušenost, nevěděli tudíž, jak bude technologie reagovat. Z rozhovorů se dají vyčíst sociální reprezentace, které mají svůj původ z kategorií, jenž jsem uvedl předchozí kapitole – *stroj*, *personifikace* a *nehmatatelný objekt*. Jelikož uživatelé nevěděli, s čím budou přesně interagovat (měli pouze zprostředkované zkušenosti), připodobnili si IVA k něčemu, co znali. Sociální reprezentace tak měli svůj původ v představách, jak uživatelé zpočátku vnímali virtuální asistenty – k čemu si je připodobňovali: stroj, personifikovaná entita, algoritmus, či zdrojový kód.

„*Když něco zvládne a zvládne to víckrát tak mám prostě důvod mu věřit a můžu mu to následně svěřovat bez strachu. Pak mu dám další věc a podobně“* (rozhovor Cyril)

„Když jsem něco hledal na internetu, tak jsem si to našel přes Siri a ona mi to řekla… No a potom sem se třeba ještě na ten internet podíval sám, jestli to tam opravdu tak je.“ (rozhovor Karel)

„Třeba budíka jsem si ověřoval. Jestli to pochopil a nastavil ho tak, jak jsem chtěl.“ (rozhovor Filip)

„Prostě jsem mu něco zadal a čekal jsem, co udělá…“ (rozhovor Erik)

„…jak bych to měla ověřovat? Když se ji ráno zeptám, jestli bude přes den pršet, tak pak uvidím, jestli mi říkala pravdu nebo ne.“ (rozhovor Aneta)

Každý z respondentů uplatnil vlastní strategii na ověřování informací, které jim jejich virtuální asistent předkládal. Z předchozí části mé práce *Počátek a prohlubování důvěry* vyplývá, že důvěra ve virtuální asistenty se formovala u respondentů postupem času, dá se říct, že jejich strategie ověřování informací byly účinné. Kdyby nebyli účinné, respondenti by si nikdy nevytvořili k nim důvěru. Z rozhovorů vyplývá, že respondenti používají výše zmínění strategie vždy, když zkouší nějakou novou funkci svého virtuálního asistenta, nebo rozšíří pole asistentova působení.

Respondenti dále uvedli, že po nějaké době přestali informace a procesy prováděné virtuálním asistentem ověřovat, nebo je neověřovali tak často. Ke kontrole práce virtuálního asistenta dochází ze strany uživatele pouze při nějakých kritických úkonech. Nelze z toho však usuzovat, že důvěra uživatelů v IVA byla zcela naplněna. Důležitý faktor, který má vliv na proces důvěry, je rozsah práce s virtuálním asistentem. Někteří respondenti by svůj vztah s virtuálním asistentem ještě více prohloubili. Nemají však k tomu v současné chvíli příležitost.

„Zatím to moc ani nejde, dát mu kontrolu, ale až budu bydlet v nějaký chytrý domácnosti, tak jí tu kontrolu budu dávat větší.“ (rozhovor Cyril)

„Tak chtěl bych samozřejmě využívat Alexu víc. Ale v současné chvíli to není možné. Ať už z finančních důvodů, technických důvodů. Řada těch věcí ještě není ani v Česku, a když jo, tak stojí neskutečné peníze.“ (rozhovor Honza)

„Ale vlastně potom, když to zkoušíš, zjistíš, že zase tolik těch situací naučených nemá. Takže vlastně můžeš zkusit většinu z nich a potom už tě tahle část moc nezajímá.“ (rozhovor Daniel)

„Určitě bych ji chtěl více využívat, ale otázkou je jak… Ta technologie se přece jen pořád vyvíjí a podle mě jsme teď fakt závislí na tom zavádění té 5G sítě, která by to mohla uspíšit.“ (rozhovor Erik)

Míru důvěry tedy ovlivňuje, kromě připodobňování si technologie k něčemu, co uživatel zná, také možnost rozsahu využívání svého virtuálního asistenta. Zde je uživatelova důvěra podporována vírou v možnosti budoucích technologií (vírou ve smyslu *confidence* dle Luhmanna [Luhmann 1979: 18–22]). Uživatel má určité znalosti o zavádění 5G sítí v České republice a vidí v tom velký potenciál pro rozmach využívání virtuálních asistentů a umělé inteligence v každodenním životě v masivním měřítku. Důvěra tak vychází z potenciálů, které mají nové technologie.

Vnímání rizika

Uvědomování si rizika plynoucího ze vzájemné interakce mezi uživatelem a jeho virtuálním asistentem je jednou z podmínek důvěry, tak jak ji definoval Luhmanna. [Luhmann 1979: 18–22] V následující části budu analyzovat a popisovat jednotlivá rizika, která si respondenti uvědomovali a rovnou je budu ukazovat v kontextu sociálních reprezentací.

Nejvýraznější obavy, které z rozhovorů vyplývaly, se týkaly nepředvídatelné chyby uvnitř asistenta, která znemožní správnou reakci na uživatelům podnět. Uvědomování si rizika vychází z faktu, že nic není dokonalé, tedy že nic není bezchybné. Ať už představa uživatelů o virtuálním asistentovi vychází z mechanického stroje, či z nedokonalého algoritmu, jedná se o objekt, který není bezchybný. Jejich sociální reprezentace si IVA vždy spojují se strojem náchylným na opotřebením, či který se nevyrovná s nenadálou událostí. Nebo ho personifikují, kdy opět dochází k asociacím

k lidské chybě. Anebo se odkazují na neschopnost lidí naprogramovat dokonalý program/algoritmus. Viz následující citace.

„Jako vždycky tam bude nějaká chyba. Může se něco stát… (…) *…ten algoritmus prostě není dokonalý. Nikdy nemůže být.“* (rozhovor Ivan)

„Nikdy se nedá myslet. Máš ten software a ten ví, jak se má chovat a určitý situace. Ale když se stane něco, co nečeká, tak nevíš, jak zareaguje a může někoho ohrozit. (…) *…je nějak naprogramovaný, ale kvůli nějaké chybě softwaru to neudělá a může ohrozit životy.“* (rozhovor Cyril)

Z citací vyplývá, že uživatelé jsou si vědomi rizik. Většinou je přirovnávají k „nějaké chybě uvnitř“. Je zde zmíněný software a algoritmus, přičemž dále v rozhovorech byly uváděny příklady s počítači. Uživatelé tak připodobňují IVA k počítačům, jakožto ke známým entitám), u kterých může vzniknout nějaká nepředvídatelná chyba. Chyba může mít svůj původ uvnitř systému, kdy respondent Ivan upozorňuje, že nelze naprogramovat bezchybný algoritmus – vždy se objeví nějaké nedokonalosti. Nebo může být chyba zapříčiněna kombinací vnějších faktorů, se kterými daná technologie nedokáže dál pracovat a vznikne chyba. Klíčové je, že chyba je nepředvídatelná a může se stát kdykoliv. Zde je důležité upozornit, že uživatelé vědí, že může nastat, a že se jí nemohou bránit.

„U všeho se může stát kravina. Někdo jede v autě, řekne ‚Siri, tell me a joke‘ a bude ten bude tak vtipný, že se ten člověk začne smát, naboural a zabije se. Jo, ale to se ti může stát i s rádiem. Stejně tak, tě GPSka může hodit jinam, než chceš.“ (rozhovor Filip)

V této citaci nejde ani tak chybu v softwaru (i když je tam také reprezentována dovětkem z GPS navigací, která může uživatele dovést na jiné místo), ale hlavně o klíčový prvek náhody, se kterým by měl uživatel počítat. Zajímavé také je, že přirovnání virtuálního asistenta v systému GPS navigací byl v rozhovorech poměrně častý. Důvodem může být, že virtuální asistenti jsou často mými respondenty využíváni k podobným činnostem, a tak je asociace s GPS zřejmá.

„Z hlediska navigace a navedete třeba špatně, kde ty mapy nebudou tak vychytaný.“ (rozhovor Daniel)

S tímto rizikem se spojené také riziko velkého spoléhání se na virtuální asistenty. Dávat virtuálnímu asistentovi moc velkou důvěru a být na něm závislý. Důvodem rizika pak může být nejenom možnost nepředvídatelné chyby. Ale také oslabení sociálních vztahů uživatele. Někteří respondenti v rozhovoru uvedli strach, že se lidé budou navzájem odcizovat.

„…právě v tom, že si na to lidé až moc zvyknou. Že ji budou nestále využívat a vlastně začnou se navzájem odcizovat. (…) *Na tohle jsem vlastně viděla film, že nějaký člověk se zamiloval do umělé*

inteligence, do robota… (…) *Já nevím, je to prostě takový divný a nemám z toho dobrý pocit.“* (rozhovor Aneta)

„…když se na ní člověk moc spolehne. Když ji bude moc důvěřovat a vlastně něco se pak stane.“ (rozhovor Bedřich)

„…že lidé budou trávit více času na mobilu a nebudou se bavit s ostatními.“ (rozhovor Jakub)

Respondentka uvádí příklad filmu[34], ve kterém se hlavní protagonista zamiluje do umělé inteligence s ženským hlasem. Film se tak vytvořil v respondentce sociální reprezentace vyvolávající obavy z možného vzájemného odcizení lidí, neboť jim umělá inteligence dokáže uspokojit veškeré jejich potřeby. Riziko tedy spočívá v přílišném propojení lidí a technologií, což by mohlo mít za následek narušení nějakého přirozeného řádu světa. Obě citace naznačují strach ze zlenivění lidí, kvůli přílišnému používání a spoléhání se na virtuální asistenty.

Z předchozí citace vyplývá, že sociální reprezentace jsou, mimo jiné, formovány kulturními faktory – například filmy. Během rozhovorů se objevilo mnoho obav pramenících ze schopností umělé inteligence, která by se v budoucnu mohla stát pro lidi hrozbou. Je pravdou, že v současné době IVA pro většinu respondentů hrozbu nepředstavuje. Bojí se však dalšího rozšíření jejich schopností.

„*Jako problém to bude, až si dokáží uvědomit sami sebe. Protože každý přemýšlející tvor si řekne, že teď sloužím já vám a už mě to nebaví a chci si dělat věci po svém a už bychom se mohli stát nástroji my.“* (rozhovor Bedřich)

„Kdyby stroje měly emoční stránku.“ (rozhovor Aneta)

„…když bude příliš chytrá, tak jestli dokáže sama myslet. A pak je těžké říci, co by si myslela.“ (rozhovor Cyril)

Respondenti poukazují na teoretickou možnost, že by schopnosti umělé inteligence byli na takové úrovni, že by představovali pro lidstvo hrozbu. Jelikož však žádná technologie v současné době takové schopnosti nemá, sociální reprezentace jsou v tomto případě tvořeny na základě zápletek v žánrů sci-fi.

Důležité riziko, které si respondenti uvědomovali, je možnost zneužití osobních údajů. Role virtuálního asistenta může být v tomto ohledu vnímána jako, jako viník – někam posílá osobní údaje uživatele, nebo může být IVA vnímán jako pouhý nástroj.

„Nesvěřil bych jí přístup do internetového bankovnictví.“ (rozhovor Cyril)

[34] *Ona* (orig. *Her*). Film z roku 2013. [česko-Slovenská filmová databáze: 2001]

„Možná nějaké zneužití údajů. Že se nám někdo naboural do domu. (…) *Teď je ale otázka, jestli to je chyba té technologie nebo někoho jiného?“* (rozhovor Honza)

„…a někdo mi může ukradnout pin, heslo, peníze. Takže z toho dle pohledu je to nebezpečné.“ (rozhovor Karel)

Možnost hackerského útoku, nebo ztráty osobních údajů si uživatelé uvědomují. Sociální reprezentace v tomto ohledu připodobňují IVA k počítači, nebo mobilnímu telefonu. Obojí je náchylné ke zneužití třetí stranou, přičemž důležitou roli hraje zabezpečení dané technologie. V tomto ohledu většinou zabezpečení IVA koresponduje se zabezpečením nosiče (stolního počítače, notebooku, mobilního telefonu, tabletu a dalších). Uvědomování si rizika se projevuje na chování uživatelů, kdy například nesvěřují svému asistentovi některé úkoly (viz rozhovor Cyril).

Často se v rozhovorech vyskytovaly obavy z mechanického poškození, nebo postupného opotřebovávání technologie. Na rozdíl od nepředpokládatelné chyby softwaru, se jedná o mechanické poškození nějaké součástky, například poškození čidla. Vlivem poškození nějaké součástky poté IVA nefunguje dobře a může být potenciálně nebezpečný.

„Tak pořád je to technika, takže stát se může všechno. (…) *Rozbít se čidlo, někde se něco zkratuje. Ale to může nastat u každého stroje.“ (rozhovor Honza)*

„U chytrých domácností. (…) Nevypne se trouba, vypne tam termostat a začne hořet.“ (rozhovor Bedřich)

Respondenti připodobňovali IVA opět ke stroji, kdy uživatelé většinou počítají s tím, že může dojít k nějakému mechanickému poškození, které má za následek nesprávné fungování technologie. Například se v rozhovoru objevilo připodobnění k autonomnímu vozidlu, kde nějaké čidlo může špatně fungovat, nezaregistruje chodce a může někoho vážně zranit. Navíc každá technologie má omezenou životnost, která se projevuje například postupným zpomalováním rychlosti fungování, nebo zvýšeným výskytem chybovosti.

V návaznosti na mechanické poruchy, uživatelé často naráželi na otázku zodpovědnosti za potenciální chyby. Kdo je v případě nějaké mechanické chyby zodpovědný, a jak bude probíhat náprava. Respondenti tak často poukazovali na nedostatečně nastavený právní rámec, který by reguloval využívání a implantaci nejenom virtuálních asistentů, ale umělé inteligence obecně do každodenního života.

„…samozřejmě otázka, zda jsme na to připravení. Já si myslím, že ne. A když se někomu něco stane, kdo bude vinen?“ (rozhovor Ivan)

Výsledky šetření a diskuze

Sociální reprezentace jako podmínka důvěry

Pro lepší pochopení zdrojů důvěry jsem se rozhodl nejprve odpovědět na podotázku, zda je důvěra podmíněná sociálními reprezentacemi o umělé inteligenci, kterou jsem si definoval v metodologické části. Až vysvětlím, jak je důvěra podmíněná sociálními reprezentacemi, bude jednoduší pochopit, jednotlivé zdroje důvěry v celkovém kontextu. V této části popíši, jak se sociální reprezentace u uživatelů virtuálních asistentů utváří a jak ovlivňují sociální realitu jednotlivce. Tím také prokážu dynamičnost sociálních reprezentací.

Sociální reprezentace, které si uživatelé vytváří na umělou inteligenci, mají významný vliv na důvěru člověka v ní. Tím podmiňují uživatelovu důvěru, jelikož dávají uživateli nějakou představu o podobě a fungování umělé inteligence, vytváří tak první zdroje důvěry, kvůli kterým je jedinec ochoten novou technologii vyzkoušet – jít do vzájemného vztahu a důvěřovat ji. Sociální reprezentace uživatelů se většinově vztahují na objekty podobné IVA (například k mobilním telefonům). Nebo na objekty zprostředkované, tedy že uživatelé nemají přímou zkušenost s technologií, ale zkušenost je pouze zprostředkovaná (připodobnění IVA k technologii ze sci-fi). Anebo že sociální reprezentace stojí na základě předchozích znalostech oboru (například fungování programování). Jedná se o sociální reprezentace, které jsou pro danou strukturu důležité. Marková a spol. je nazývají sociální reprezentace jádra[35]. Tyto reprezentace formují sociální vědomí nějaké struktury o něčem neznámém. [Marková a spol. 1998: 805–827] V mém výzkumu se jedná o reprezentace vycházející z kategorií *strojovost IVA, personifikace IVA a nehmatatelnost IVA*, které na počátku interakce uživatelů s asistentem dali počáteční impuls virtuálnímu asistentovi důvěřovat – to také dokazuje podmíněnost počáteční důvěry sociálními reprezentacemi.

V rozhovorech moji respondenti uvedli, že do vzájemného vztahu s IVA vstupovali s určitou představou založenou na zprostředkovaných zkušenostech. Prostřednictvím reklam, videí a článků na internetu, či prostřednictvím svého sociálního okolí uživatelé měli určitou představu, jak IVA vypadá. V první fázi tvorby sociálních reprezentací dochází k *ukotvení* informací, kdy se virtuální asistent stává součástí sociální reality uživatele. [Moscovici 1988: 211–250] Uživatel vnímá IVA primárně tak, jak ho poznal – mobilní telefon, součást telefonu, funkce v telefonu, nebo krabička, která ovládá domácnost. V tuto chvíli má uživatelova představa o IVA jasnou formu, ale stále nemá

[35] Jádra a periferie se spolu asymetricky vyměňují. Předpokládá se, že jádra se mohou měnit pomaleji než periferie. [Marková a spol. 1998: 825] Vzájemná asymetrie vytváří dynamiku sociálních reprezentací, viz kapitola: *Další pojetí teorie sociálních reprezentací* str. 13.

obsah. Uživatel už ví, jak vypadá, ale stále neví, co od virtuálního asistenta očekávat. Proto nastává druhá fáze tvorby sociálních reprezentací – *objektivizace*. V této fázi uživatel aktivně přiřazuje právě ukotvené informace do svého vědomí k jemu už známým kategoriím. V praxi to znamená, že uživatel přiřazuje objektu takové vlastnosti, o kterých se domnívá, že jimi objekt disponuje. Nejedná se však o z nahodilé přiřazování vlastností, ale o systematický proces, kdy uživatel sumarizuje vědění o konkrétní věci. [Moscovici 1988: 211–250] Uživatelé tak virtuálním asistentům přiřazovali vlastnosti, či funkce, které má například jejich mobilní telefon, nebo počítač. Díky tomu může uživatel s pojmem inteligentní virtuální asistent aktivně nakládat. Reprezentace jsou soustavně ukotvovány a transformovány novými informacemi a prožitky, což ovlivňuje jak nové sociální reprezentace, tak i ty staré. Následkem je postupná změna vnímání okolní reality a vnímání jedince či celé sociální struktury. [Höijer 2011: 7–12]

Také se potvrdila dynamičnost sociálních reprezentací, respektive její neustálé proměny v čase. To opět potvrzuje poznatky Markové a spol., která uvádí, že sociální reprezentace a jejich struktury se neustále mění. To má za následek proměňování sociálních praktik a změnu sociální reality. [Marková a spol. 1998: 825] V případě mých respondentů se proměna sociálních reprezentací v čase projevuje jiným vnímáním virtuálních asistentů a tím změnou míry důvěry. Za proměňováním jejich reality, tedy za proměnami vnímání jejich virtuálního asistenta, stojí především dynamika sociálních reprezentací. A přestože má dynamičnost sociálních reprezentací u každého uživatele jinou intenzitu, v určité míře se objevuje u všech mých respondentů. Marková dále uvádí, že k proměnám sociálních reprezentací dochází kvůli asymetrické výměně reprezentací z jádra a periferie. [Marková a spol. 1998: 825] Dynamika proměny sociální reality u uživatelů IVA naznačuje určitou asymetrii mezi sociálními reprezentacemi (viz příklad s respondentkou Anetou).

Vlivem nových prožitků dochází k vytváření sociálních reprezentací, které mohou být v rozporu se starými reprezentacemi. Respondentka Aneta uvedla, že virtuálního asistenta vnímá především jako stroj. Na základě výzkumného rozhovoru jsem zjistil, že proces přebírání a vysvětlení pojmu IVA, jakožto neznámé entity, probíhal velmi podobně, jako jsem popsal výše. Když se však virtuální asistent aktivně zeptal respondentky Anety *„Jak se dneska máte?“,* vyvolalo to v respondentce novou reprezentaci. Zážitek byl natolik silný, že došlo k ukotvení něčeho, co bych nazval aktivní komunikací.[36] Respondentka si ukotvila zážitek, kdy ji stroj aktivně oslovil a zeptal se ji na nesouvisející věc, s aktuální činností. Ve druhém kroku došlo k objektivizaci, kdy prožitek mluvícího

[36] Respondentka uvedla, že se jí Siri zeptala sama od sebe při interakci, která se týkala nějaké každodenní činnosti na virtuálním asistentovi.

stoje přiřadila pro ni k nejbližší kategorii – mluvící roboti ze sci-fi filmů[37]. Tak přiřadila svému virtuálnímu asistentovi nové vlastnosti – komunikativní, inteligentní, potenciálně nebezpečný. V tomto případě ovlivnily sociální reprezentace důvěru uživatelky negativně. To značí existenci dynamiky sociálních reprezentací, která má vliv na proměny vnímání reality uživatele.

V analýze se zaměřuji primárně na zkoumání důvěry, její aspekty a proměnu v čase. Proto jsem nevěnoval takovou pozornost kategorizování různých sociálních reprezentací. Na základě zjištěných dat mohu potvrdit hierarchičnost sociálních reprezentací, o které mluví Marková a spol. [Marková a spol. 1998: 805–827] Sociální reprezentace si nejsou rovny, přičemž reprezentace tvořeny sociální strukturou mají významnější vliv, než reprezentace periferie, který získává uživatel při interakci s IVA. Také mohu říci, že většina analyzovaný reprezentací spadá do jedné ze tří kategorií – *strojovost IVA, personifikace IVA a nehmatatelnost IVA*. Sociální reprezentace jádra jsou pak přímo spojená s jednou z těchto kategorií, zatímco jejich nadstavby (v analýze charakteristické znaky) jsou už sociální reprezentace periferie. Například uvažování o umělé inteligenci jako o stroji, je sociální reprezentace jádra, ale přisuzování ji možnosti softwarové chyby je již periferie. Stejně tak sociální reprezentace spojené personifikací jsou v jádru, konkrétní připodobnění a vlastnosti jako otrok, asistent, či kolega jsou už sociální reprezentace jádra. Na druhou stranu, vzhledem k nastavení designu výzkumu nemohu strukturálně rozdělit sociální reprezentace na reprezentace jádra a periferie, což také ani nebylo cílem výzkumu. Nicméně právě strukturální rozdělení sociálních reprezentací ohledně umělé inteligence je zajímavé téma, které může a mělo být předmětem dalšího zkoumání.

Zdroje důvěry v umělou inteligenci

Během své analýzy jsem objevil čtyři základní zdroje, které ovlivňují důvěru uživatele ve virtuálního asistenta. Dva zdroje důvěry považuji za stabilní – *neutralita* a *víra v budoucnost* a dva za podmíněné *naplnění očekávání* a *blízkost*. V následující části tedy podrobně zodpovím moji výzkumnou otázku, jaké jsou zdroje důvěry. Jednotlivé zdroje charakterizuji a poukáži na jejich odlišnosti a následně popíši, jak se proměňují v čase.

Zdroje důvěry, které se mi na základě analýzy podařilo určit, nemají stejný základ. Některé zdroje důvěry jsou stabilní a vychází z jádra sociálních reprezentací struktur (viz předchozí část). Jedná se o zprostředkované zkušenosti, které vychází ze sociálních struktur. Nepřímé zkušenosti v jedinci vyvolávají pocit sociální důvěry, který má za následek všeobecné přijetí technologie. Jedná se o tak

[37] Respondentka v rozhovoru uvedla filmy Ona a Terminátor.

zvanou kognitivní cestu k důvěře. [Liu, Yang, Xu 2019: 326–341] Ty slouží jako první motivace člověka mít ochotu jít do vzájemné interakce s umělou inteligencí. Jedná se o stabilní důvěru, kterou do určité míry pociťuje každý jednotlivec. Hoff a Bashir píší o tzv. dispoziční důvěře, která je také stabilní a má svůj původ v jednotlivci (věk, pohlaví, osobnost, kulutra). [Hoff, Bashir: 2014] To do jisté míry potvrzuje i moje výstupy. Stabilními zdroji důvěry je *pocit neutrality* a *víra v budoucnost*. Dle Lewickiho a Benkerové se jedná o 1. fázi důvěry, kdy člověk něčemu důvěřuje, na základě nedostatku informací. [Lewicki, Bunker 1995: 133–173]

Naopak zdroje důvěry *blízkost* a *naplnění očekávání* jsou podmíněné osobní zkušeností a nejsou tak příliš závislé na sdílené zkušenosti struktur. Na základě kumulované zkušenosti si uživatel vytváří nový pohled na virtuálního asistenta, který má vliv na míru důvěry. Liu, Yang a Xu to nazývají afektivní cestou, která je právě založená na osobní zkušenosti a pro jedince má tak důležitější význam. To znamená, že má na uživatele i silnější vliv. [Liu, Yang, Xu 2019: 326–341] Konkrétní příklady uvedu u jednotlivých zdrojů důvěry.

Prvním zdrojem důvěry je *neutralita*. Umělá inteligence vyvolává v uživatelích zdání neutrality a nestrannosti. Boyd a Crawford píší, že lidi mají tendenci přistupovat k technologiím neutrálně, aniž by si byli vědomi, že každá technologie byla vytvořena za nějakým účelem. Lidé mají tendenci přehlížet některá fakta o technologiích, jako například kdo je vyrobil, za jakým účelem byly vyrobeny, nebo že je musí někdo udržovat (myšleno ve smyslu updatovat). [Boyd, Crawford 2012: 662–679] Uživatelé tak nekalkulují s tím, že by je jejich virtuální asistent chtěl ze své vůle nějakým způsobem poškodit, nebo jim jinak ublížit. Automatizace nemá vlastní úmysly a ze své vůle nemůže uživatelům ublížit. [Lee, See 2004: 65] Jinými slovy, existuje zde počáteční důvěra, protože zde není riziko určité formy „zrady“. Frederiksen o tom mluví v rámci dimenze mezilidských vztahů, kdy mezi lidmi automaticky existuje riziko, že ten druhý je podvodník. [Frederiksen 2012: 733–750] U technologie je toto riziko uživatelem silně potlačeno. To také odpovídá konceptu slabého induktivního vědění od Giddense, kdy je člověk do jisté míry odkázán technologii věřit, neboť nemá dostatečné znalosti. [Giddens 2003: 33–39] Vnímání neutrality virtuálního asistenta tak je stabilním zdrojem důvěry pro uživatele, který je možné nalézt v přibližně stejné míře napříč celou sociální strukturou.

Druhým zdrojem důvěry je *víra v budoucnost*. Jedná se o zdroj důvěry, který je stejně jako zdání neutrality stabilním napříč sociální strukturou. Spočívá v tom, že uživatel natolik věří možnostem a potenciálu umělé inteligence, že tato víra neustále posiluje jeho důvěru v UI a tím se stává zdrojem důvěry. Umělá inteligence je uživateli vnímána jako role/instituce, které jsou připisovány expertní vlastnosti – lepší než člověk, menší chybovost, menší náklady na provoz a podobně. To koresponduje

Giddensovým pojmem důvěra v *expertní systémy*. [Giddens 2003: 75–101] Důvěra je navíc posilována vědecko-populárními proudy, které vyzdvihují určité vlastnosti umělé inteligence a lidé tak mají tendenci vlastnosti umělé inteligence přeceňovat. Ke stejnému závěru došli i Berling a Bueger [2017: 1–10]. *Víra v budoucnost* je vírou lidí v potenciál technologií, které jednoho dne budou téměř dokonalé a budou umět všechno. Jedná se o slepou víru v něco ve smyslu *confidence*, dle Luhmannovi teorie důvěry. [Luhmann 1979: 18–22] Jedinec tak upřednostňuje nějaké aspekty technologie na úkor potenciálních rizik. Na základě osobních zkušeností může být tento zdroj důvěry u uživatele posilován nebo oslabován. V určité míře je však pevnou součástí sociální struktury.

Třetím zdrojem důvěry je *naplnění očekávání*. Tím je myšleno, že virtuální asistent plní zadané úkoly správě, dle očekávání, a tím uspokojuje požadavky uživatele, který mu na základě toho svěří další činnost. Tím ve výsledku dochází k prohlubování důvěry. Jedná se o zdroj důvěry, který je závislý na kumulovaných zkušenostech uživatele. Během rozhovorů bylo jasně patrno, jak splněná očekávání postupem času proměňovala uživatelův pohled na virtuálního asistenta a tím i proměňovala podstatu samotné důvěry. Potvrzuje to tedy sílu osobních zkušeností, která je větší než síla zkušeností zprostředkovaných, jak uvádí Liu, Yang a Xu. [Liu, Yang, Xu 2019: 326–341] Také to odpovídá 2 fázi důvěry podle Lawického a Bunkera, kdy je důvěra založená na znalostech – v tomto ohledu tedy na zkušenostech pro uživatele důležitější a na základě toho dochází k prohlubování důvěry. [Lewicky, Bunker 1995: 133–173]

Zcela zásadní vliv hraje v kontextu naplnění očekávání *temporalita*, tedy doba, po kterou dochází k naplňování očekávání. Z rozhovorů jasně vyplývá, že čím více má člověk s danou technologií zkušeností, tím více má s ní prožitků a tím spíše jí důvěřuje. Také se v průběhu času proměňují úkoly, které virtuální asistent plní. Čím širší je škála činností, které virtuální asistent vykonával, nebo vykonává, tím více má s ní uživatel zkušeností (kladných i záporných). Zná tak limity svého asistenta a dokáže tak přesně určit, v čem mu může důvěřovat a v čem zatím ne. Purwanto, Kuswandi a Fatmah uvádí, že lidé budou chtít s virtuálními asistenty více komunikovat a asistenti se tak proto musejí neustále vyvíjet dopředu. [Purwanto, Kuswandi a Fatma 2020: 72] V rámci mé práce je potřeba k tomuto tvrzení ještě doplnit, že pokud nebude docházet k dostatečnému vývoji schopností virtuálních asistentů, uživatelé tak nebudou moci rozvíjet své zkušenosti s IVA a tím i nebude docházet k posouvání důvěry ve virtuální asistenty. Například, pokud by se nyní přestali schopnosti virtuálních asistentů rozvíjet, za nějakou dobu by je uživatelé začali považovat za zastaralé a neschopné, neboť by již neodpovídali jejich současným požadavkům. Dalším zjištěním mé práce tedy je, že *termoralita* má významný vliv na plnění očekávání, přičemž *naplnění očekávání* je důležitý zdroj důvěry. Mnoho jiných autorů se však temporalitou ve vztahu k důvěře a k plnění očekávání vůbec nezabývá, například [Rotter: 1967; Johns: 1996; Lee, See:2004; Liu, Yang, Xu: 2019] Tím se také tato práce

od těchto autorů liší. Z mého šetření jasně vyplývá, že temporalita je důležitý aspekt zkoumání důvěry v UI ve vztahu ke zkušenostem, který i do budoucna nabízí velký potenciál pro sociologická zkoumání. Naopak potvrzuji zjištění Hoffa a Bashira, kteří došli ve svém bádání ke stejnému zjištění, tedy, že tzv. *naučená důvěra*[38] je jednou z možný vrstev důvěry v technologii. [Hoff, Bashir: 2014]

Čtvrtým faktorem důvěry je *blízkost*. Vzájemnou interakcí člověka a virtuálního asistenta dochází k vytváření vztahu, který má vliv na míru důvěry. Frederiksen mluví o dimenzi blízkosti v rámci mezilidských vztahů, kdy emocionální blízkost vytváří prostředí pro důvěrný vztah. [Frederiksen 2012: 733–750] Frederiksenova slabá blízkost se několikrát objevila v mých rozhovorech, kdy respondenti přirovnávali IVA k asistentovi, či kolegovi. Stejné vlastnosti jako má reálný kolega v práci, očekává uživatel i od svého virtuálního asistenta, a to definuje jejich vzájemný vztah a důvěru. Mohlo by se zdát, že *blízkost* je tedy pouhým produktem naplnění očekávání. V samotných počátcích vztahu tomu tak pravděpodobně je. Po nějaké době si však uživatel může vybudovat s IVA určitý vztah, kdy už *blízkost* působí jako emocionální zdroj důvěry. V několika citacích z rozhovorů je patrno, že uživatelé pociťují určité emoce při komunikaci s virtuálním asistentem. Například respondent Bedřich, který uvedl: *„Ten pocit, že virtuální asistent může být tvůj mobil, že na tebe mluví mobil, má ženský hlas a je to takový příjemný."* Nebo respondent Jakub, který během celého rozhovoru mluvil o IVA jako o živém asistentovi. V obou případech se dá mluvit o určitém emocionálním poutu respondentů s IVA, který se stal jejich zdrojem důvěry. Frederiksen [Frederiksen 2012: 733–750] však ve své práci nepočítá s možností blízkosti mezi jinými subjekty než dvě živé bytosti. V rámci mé práce jsem však došel k závěru, že *blízkost* může být i mezi člověkem a jeho virtuálním asistentem, a že se také může v průběhu času prohlubovat. Na základě získaných dat není v mých silách určit, kterých ze tří Frederiksenových dimenzí blízkosti může vztah mezi člověkem a UI nabývat, ale určitě by to také mohlo být námětem nějakého dalšího výzkumného šetření, které by navazovalo na moji práci. Bez zajímavosti také není, že i zde se opět objevuje faktor temporality, který má důležitý vliv na proměnu důvěry.

Blízkost má však význam i ve smyslu fyzické blízkosti. IVA je v mobilním telefonu, nebo v počítači, popřípadě jako ovladač domácnosti, který má uživatel neustále na blízku. Mirilello uvádí, že mnoho lidí má mobilní telefon neustále při sobě. Stává se pro uživatele důležitým a neradi ho někomu svěřují. Naopak si ho lidé často hlídají a tráví s ním mnoho času. [Mirilello a spol. 2013: 89–95] Virtuální asistenti například nabízí možnosti ovládání prostřednictvím hlasových povelů. Stává se tak prostředníkem mezi uživatelem a například jeho mobilním telefonem. Tím, že virtuální asistent neustále reaguje na hlasové povely, vyvolává v uživateli pocit, že je neustále na blízku a tím se mezi

[38] Důvěra založená na získaných zkušenostech.

nimi opět buduje důvěrný vztah. Příkladem může být respondent Daniel, který prostřednictvím IVA provádí na telefonu mnoho různých operací – nastavení, vyhledávání, přepínání písniček apod.. Jeho virtuální asistent se tak stal přímým spojením mezi uživatelem a jeho telefonem. Postupem času se tak buduje vztah, kdy oba subjekty jsou na sebe zvyklý a vědí, co očekávat. Výsledkem je, že uživatel ovládá svůj telefon spíše prostřednictvím virtuálního asistenta než napřímo.

Důvěra u uživatelů virtuálních asistentů se proměňuje v čase. Existují zdroje důvěry, které jsou fixně dány sociální strukturou, kdy se důvěra opírá o expertní vědění a nedostatek znalostí. Ty následně rozšiřují zdroje důvěry, které jsou něčím podmíněné, tedy jsou založeny již na osobních zkušenostech. To odpovídá vývoji důvěry podle tří fází Lawického a Bunkera. [Lewicky, Bunker: 1995] Na druhou stranu se mi na základě získaných dat nepovedlo prokázat přítomnost 3. fáze důvěry, tedy důvěry jedince ve virtuálního asistenta založené na sociální identifikaci. Je otázkou, zda se člověk může někdy v budoucnu plnohodnotně identifikovat s virtuálním asistentem. Dle dostupných zjištěních v této práci to však za možné nepovažuji. Důvodem může například být silná hierarchická asymetrie vztahu mezi uživatelem a jeho virtuálním asistentem, která v současnosti vzájemnou identifikaci vylučuje. Dalším důvodem mohou být schopnosti virtuálního asistenta, které buď vysoce převyšují schopnosti člověka, nebo jich naopak nedosahují. V obou případech tak dochází k výrazné dovednostní nerovnováze, která nenaznačuje možnost sociální identifikace, kterážto je založená na podobnosti jedinců. Je tedy otázkou na kolik je pojetí a vývoj důvěry dle [Lewicky, Bunker: 1995] pro studium důvěry uživatelů k umělé inteligenci relevantní.

Vnímání rizika uživatelů virtuálních asistentů

Podle Niklase Luhmanna je uvědomování si rizik, plynoucích ze vzájemné interakce, základním kamenem důvěry. Bez uvědomování si rizik se nejedná o důvěru (trust) ale důvěřivost neboli slepou víru v něco (confidence). [Luhmann 1979: 18–22] Proto je sledování rizik, která si uživatelé uvědomují, důležitou součástí mé práce. Respondenti během výzkumných rozhovorů mluvili o šesti rizicích: *nepředpokládaná chyba softwaru*, *mechanická chyba*, *zneužití IVA třetí stranou*, *právní rizika*, *strach z odcizení* a *ovládnutí technologiemi.* V následující části jednotlivá rizika popíši a ukáži, jaký mají vliv na uživatelovu důvěru. Také navrhnu další možné výzkumy rizik, kterými by se mohly zabývat další práce.

Nepředpokládaná chyba softwaru. Riziko, kterého se nějakým způsobem dotkl téměř každý z respondentů. Jedná se o riziko nesprávného chování virtuálního asistenta na základě nepředpokládané chyby v softwaru, či nesprávně nastaveného algoritmu, který měl za následek asistentovo nesprávné pochopení situace a vyvození mylného závěru. Uživatelé připodobňují IVA

ke stroji, člověku, či algoritmu. Všechna tato přirovnání jsou však k entitám, které nepracují bezchybně. Uvažování o nepředvídatelných rizicích je založeno na zkušenostech uživatele s entitami, ke kterým technologii připodobňuje. Považují-li uživatelé virtuálního asistenta za stroj, je logické očekávat podobnou chybovost, jako stroj. Stejně tak se dá uvažovat u přirovnání k člověku či algoritmu. Přesto se však projevila iluze *důvěryhodnosti* a *spolehlivosti stroje*, kterou popisuje Madhavan a Wiegmann. [Madhavan, Wiegmann: 2007] Uživatelé v rozhovorech několikrát uvedli, že nějakou činnost[39] by raději nechali provést virtuálního asistenta než někoho jiného. Zde se prolíná *důvěryhodnost* technologie s iluzí větší *spolehlivosti stroje*. Na druhou stranu Madhavanův a Wiegmannův třetí pojem – *zdroj diagnostických operací* v tomto případě neplatí. Uživatelé jsou podle autorů mnohem vnímavější k chybám strojů než k chybám lidí, což plyne z přesvědčení, že stroje jsou spolehlivější. [Madhavan, Wiegmann 2007: 281–288] Z rozhovorů však vyplynulo, že uživatelé spíše chyby virtuálních asistentů omlouvají. Spolehlivost virtuálních asistentů většinou považují za lepší než spolehlivost u jiných lidí, ale přesto jsou si vědomi rizika nepředpokládané chyby – softwaru i hardwaru. S vědomím možné rizikovosti tak chyby u IVA přehlížejí a omlouvají. Je tedy zajímavé, že v tomto případě Madhavanův a Wiegmannův iluzorní pohled uživatelů neplatí.

Mechanická chyba. S logikou předchozího rizika nepředpokládané chyby softwaru souvisí také riziko mechanické chyby. Tím jsou myšleny obavy z postupného opotřebování, zpomalování technologie. Strach z rozbitých, znečištěných, či jinak nefunkčních senzorů, které by měli za následek nesprávné fungování asistenta. I zde připodobňují uživatelé virtuálního asistenta ke strojům, které se mohou mechanicky poškodit. Zatímco u předchozího bodu se jednalo o chybu spíše nepředpokládanou, v případě mechanického poškození, nebo opotřebení s tím uživatelé do jisté míry počítají. Jedná se tedy o chybu, pro uživatele, očekávatelnou. Uživatelé opět vycházejí ze svých zkušeností. Nicméně i zde měli respondenti tendence mechanické chyby omlouvat. Nejčastějším argumentem bylo přirovnání k jakémukoliv jinému stroji, kde může podobná situace nastat také. Přístup uživatelů IVA je v případě rizika mechanické chyby podobný, jako jejich přístup v předchozím bodě. Což je opět v rozporu s třetím iluzorním pohledem Madhavana a Wiegmanna, kteří uvádějí, že lidé bývají vnímavější k chybám strojů, neboť u stroje se potenciální chyba nepředpokládá. [Madhavan, Wiegmann 2007: 281–288]

Zneužití IVA třetí stranou. Respondenti v rozhovorech často uváděli riziko možnosti zneužití osobních údajů, či hackerského útoku. Jelikož se jedná o podobná rizika, jejichž následky mohou být podobné, sloučil jsem obě rizika do jednoho. Nicméně se vnímání rizika zneužití u uživatelů liší a dělí na dvě skupiny. V prvním případě je zneužití osobních údajů vážným porušením důvěry ve

[39] V rozhovorech uváděli: vyhledávání na internetu, nalezení nejrychlejší cesty, zjišťování aktuálních informací.

virtuálního asistenta a je důvodem pro ukončení vzájemné spolupráce. Na druhou stranu někteří respondenti nevnímají zneužití osobních údajů jako vinu virtuálního asistenta, ale jako vinu třetí osoby. IVA je pro ně pouhým zneužitelným nástrojem, který nemá moc se útoku nějakým způsobem bránit. Pro tuto skupinu respondentů by tak zneužití osobních údajů, či hackerský útok, nebyl důvodem ke ztrátě důvěry ve virtuálního asistenta, ale důvodem k prohloubení nedůvěry v lidi a zabezpečovací systémy. Purwanto, Kuswandi a Fatmah uváží, že zneužití osobních údajů je jedním z největších rizik při vzájemné interakci uživatele a virtuálního asistenta. Zároveň si klade otázku, co by přimělo člověka přestat využívat svého IVA. [Purwanto, Kuswandi, Fatmah 2020: 70–72] Zcizení osobních údajů je pro některé respondenty vážným rizikem a jakékoliv zneužití by pak pro ně znamenalo oslabení důvěry a ukončení využívání virtuálního asistenta. Z rozhovorů však není zcela jasné, zda se to týká pouze onoho konkrétního asistenta, nebo zda by uživatel ztratil důvěru k IVA obecně a nikdy žádného už by už nevyužíval. Dle odpovědí respondentů nad tím ani tak hluboce nepřemýšleli. Někteří respondenti uvedli, že se bojí ztráty osobních údajů, ale následně dodali, že jim nevadí poskytovat své osobní údaje virtuálnímu asistentovi, neboť díky nim může IVA lépe dělat svoji práci. Stejně tak by někteří respondenti byli ochotni poskytnout i více osobních údajů za účelem lepší vzájemné spolupráce. Boyd a Crawford popisují lidskou neutralitu k zpracovávání osobních údajů technologiemi jako mylnou, neboť každý algoritmus nese stopu „lidského úmyslu“. Data, která IVA zpracovává, mohou být upravována, transformována, či ořezávána a tím i měněn jejich smysl. [Boyd, Crawford 2012: 662–679] Na základě zjištění také můžu potvrdit předpoklad Belangera a Xu, že někteří uživatelé nakládají se svými osobními údaji nesystematicky a iracionálně. [Belanger, Xu 2015: 575] To se projevuje značnou obavou ze ztráty osobních údajů některých uživatelů, ale na druhou stranu jim nevadí své osobní údaje poskytovat v dosavadním nebo i větším měřítku. Proto se domnívám, že uvažování uživatelů nad poskytováním osobních údajů a riziky s tím spojených je zajímavým námětem pro další výzkum, který by mohl přinést mnohá zajímavá zjištění.

Právní rizika. Tento pojem pod sebou sdružuje širokou paletu obav z nepřipravenosti jednotlivců i společností na silnou provázanost technologie a lidstva. Respondenti se zde odvolávají na právní rámce a regulatorní systémy, které jsou spojené s masivním rozšířením umělé inteligence. Mezi nejčastějšími příklady byla uváděna trestní odpovědnost. Když umělá inteligence někomu omylem způsobí újmu na zdraví, nebo na životě, kdo je za to zodpovědný? Umělou inteligenci postihnout nelze. Je možné postihnout jejího vývojáře, když víme, že nejde vytvořit dokonalý algoritmus, který bude pracovat bez chyby. Tím se dostávám k druhé části tohoto bodu, kdy respondenti často upozorňovali na nedostatečný, nebo nejasný systém zabezpečení umělé inteligence. To do jisté míry také koresponduje s předchozím bodem, kdy se lidé bojí zcizení jejich osobních údajů. Je možné, aby nedokonalý člověk vytvořil dokonalý zabezpečovací systém? Tento bod upozorňuje na širokospektrý multidisciplinární problém, který vyžaduje rozsáhlou diskuzi napříč jednotlivými

vědními disciplínami. Jelikož se tato práce primárně soustředí na zkoumání důvěry a vlivů rizik na důvěru v UI, není mým cílem zde popisovat rozsáhlé diskuze a argumenty, které již byly řečeny v jiných akademických textech [French: 1990; Boström: 2003; Boström, Yudkowsky: 2014; Belanger, Xu: 2015].

Strach z odcizení. Někteří respondenti pociťují strach, že kvůli schopnostem virtuálních asistentů s nimi budou lidé chtít ještě více interagovat. Obavy pramení ze zkušenosti, kdy stále více lidí komunikuje se světem prostřednictví technologií na úkor přímé interakce s lidmi. Tato obava koresponduje s nálezy Mirilello a spol., kdy upozorňuje, že lidský čas předurčený ke komunikaci s ostatními je omezený. Lidé pomocí komunikačních technologií mohou nyní komunikovat častěji a intenzivněji než dříve, mohou však obsáhnout mnohem menší sociální prostor. Člověk tak rozděluje svůj komunikační čas mezi menší skupinu lidí a tím si oslabuje své vazby na širší sociální okolí. [Mirilello a spol. 2013: 89–95] Tyto obavy jsou samozřejmě opodstatněné, otázkou však je, do jaké míry může IVA ovlivnit čas uživatele například na mobilním zařízení. Mnoho uživatelů tráví hodně času na svém mobilním zařízení, aniž by IVA využívali. Navíc respondenti v rozhovorech často uvádí, že využívají virtuálního asistenta, protože je s ním práce jednoduší a rychlejší. Z logiky věci by tedy mělo vyplývat, že IVA čas strávený na mobilu uživatelem spíše zkracuje. V tomto případě se z mé strany jedná o pouhý dohad. Nicméně vliv virtuálního asistenta na čas strávený uživatelem na jeho mobilním zařízení je, dle mého názoru, opět zajímavý námět ke zkoumání, který by do dané problematiky vnesl trochu více světla.

Ovládnutí technologiemi. Většina respondentů mluvila o nadvládě robotů, či nadvládě nějaké pokročilé technologie jako o něčem, co v současné době není možné (z důvodu nedostatečné vyspělosti technologie). Na druhou stranu v průběhu většiny rozhovorů si respondenti uvědomovali možné potenciální ohrožení lidské existence. Důvodem je dlouholetá tradice sci-fi kultury, která od 60. let minulého století dramaticky vykresluje budoucnost a zaměřuje se na pocity nejistoty a obav z technologií a budoucnosti obecně. [Maříková, Petrusek, Vodáková a spol. 1996: 969–970]. Navíc Peggs[40] upozorňuje, že lidé chtějí být vždy hierarchicky nadřazeným druhem na planetě, přičemž nechtějí přijít o privilegia, která z toho plynou. Proto lidé podvědomě nechtějí dát stejnou moc (práva) nějaké jiné entitě (v tomto případě UI). [Peggs 2013: 591–606] Z toho plyne podřízený vztah mezi uživatelem a jeho IVA, který je patrný v každém rozhovoru.

[40] Peggs popisuje vzájemnou trvalou nerovnost mezi domácími mazlíčky a jejich pány. Přestože je v mnoha případech pes, či kočka brána jako člen domácnosti, nikdy nebude mít stejná práva, jako lidé. Dále ukazuje, že i přístup sociologického bádání je v tomto smysle nevyrovnaný a silně protěžující lidskou populaci. [Peggs 2013: 591–606]

Mezi respondenty se objevují také odpovědi, ve kterých si uživatelé nejsou vědomi žádných rizik. Uživatelé tak slepě důvěřují technologii, což Luhmann za důvěru nepovažuje. [Luhmann 1979: 18–22] Na druhou stranu uvažování, které se objevuje v citacích, koresponduje s Giddensovým pojetím důvěry v expertní systémy [Giddens 2003: 75–101], kdy uživatelé nemají dostatek potřebných vědomostí, aby si mohli udělat vlastní názor. Proto se rozhodli důvěřovat vývojářům a distributorům, a tím i důvěřovat správnému nastavení a naprogramování IVA (viz část *Zdroje důvěry v umělou inteligenci*).

Celkově si však většina respondentů nějaká rizika uvědomovala, což ovlivňovalo jejich přístup k virtuálnímu asistentovi. Při interakci s asistentem tak zvažovali různá rizika a jejich dopady. Liu, Yang a Xu upozorňují, že rizika jsou predikátorem přijetí nějaké technologie, nicméně pokud se jedná o zprostředkovná rizika, jejich síla není tak velká. [Liu, Yang, Xu 2019: 326–341] To do jisté míry může tato práce potvrdit, nebo strach z mechanické chyby má u uživatelů silný základ, neboť s ní má mnoho uživatelů již nějakou skutečnost. Uživatelé tak počítají s postupným zpomalováním reakcí virtuálního asistenta, nebo s poškozením nějaké jeho části. Naopak v případě zneužití osobních údajů, uživatelé nijak neměnili svoje chování vůči virtuálnímu asistentovi, neboť hrozbu zneužití znají pouze zprostředkovaně a nemá pro ně takovou váhu.

Faktory ovlivňující důvěru uživatelů v umělou inteligenci

Sociální faktor

Důležitým sociálním faktorem, který má vliv na tvorbu sociálních reprezentací u uživatelů IVA, jsou zprostředkované a osobní zkušenosti. Uživatelé prostřednictvím médií získávají zprostředkovné zkušenosti ve formě vědeckých, či vědecko-populárních pohledů na nové technologie, které následně zásadně ovlivňují jejich pohled na virtuální asistenty. [Berling a Bueger 2017: 1–10] IVA v nich bývají prezentováni jako nejmodernější technologie, téměř neomezených možností. [Hengstler, Enkel a Duelli: 2016: 105–114] Osobní zkušenosti postupně upravují původní pohled na virtuální asistenty a tím vytváří nové sociální reprezentace. Příkladem je přirovnání virtuálního asistenta v otrokovi, či zaměstnanci, kdy vlivem osobních zkušeností dochází k přirovnání technologie k živé entitě.

Využívání IVA je často spojováno se značkou, která virtuálního asistenta nabízí (Siry – Apple, Google assistant – Google, Alexa – Amazon apod.). Vlastnění virtuálního asistenta se znamená v některých rozhovorech pro uživatele vlastnění socio-ekonomického statku a znamení prestiže. Podobně jako vlastnictví iPhonu může v naznačovat určitý socio-ekonomický status, samotné

vlastnění a využívání virtuálního asistenta je spojováno mezi uživateli s určitou mírou prestiže. Tato prestiž se následně promítá do tvorby sociální reprezentací, kdy uživatelé vnímají používání virtuálních asistentů jako určitá známku prestiže, nebo vlastnictví inteligentní domácnosti jako socio-ekonomický statek.

Ekonomický faktor

Sociální reprezentace uživatelů byly silně ovlivněny ekonomickými faktory. Zejména v oblasti potenciálu umělé inteligence. Často mluvili o schopnostech současných technologií, které se vyznačují svojí rychlostí, efektivitou a ekonomickou úsporou. Kvůli technologii se mohou věnovat jiným činnostem, nebo jim činnost trvá minimální čas. Uživatelé často vnímali IVA jako profesionálního asistenta, či dokonce kolegu v práci, který jim uspořil čas, či zefektivnil práci. Respondenti často umělou inteligenci připodobňovali strojům v automobilkách, či jiných zařízení, kde stroje vykonávají různé práce. To s sebou přináší úsporu nákladů za pracovní sílu a zároveň vysokou míru efektivnosti. To se také projevovalo v kódu průmyslová výroba, která je s novými technologiemi často spojována a projevuje se i v sociálních reprezentacích uživatelů.

Dalším faktor, který si respondenti uvědomovali, je různá forma zneužití jejich osobních údajů, či hackerský útok, prostřednictvím virtuálního asistenta. Uživatelé kalkulují s možností rizika zneužití jejich údajů, který by měl za následek ohrožení jejich ekonomických statků. Toto uvažování se tak projevuje i v tvorbách sociálních reprezentací uživatelů. Ze stejného důvodu také uživatelé kalkulují s mírou důvěry, kterou jsou ochotni IVA dát. Častým faktorem byla finanční ztráta v případě asistentova selhání. Tento faktor se tak projevoval v sociálních reprezentacích zejména v situacích, kdy uživatel přemýšlí, zda svěří svému IVA kontrolu nad něčím.

Kulturní faktor

Během celé práce se ukazoval jako silný faktor sociálních reprezentací tradice sci-fi. Kultura sci-fi je projevuje rozvinutím imaginace, vyjádřením specifického životní filosofie a dramatickým vykreslením vize budoucí společnosti, vztahů mezi jedinci a jejich způsobu života. Často se v ní objevuje princip obav a hypertrofie technické dimenze civilizace. Charakteristické jsou pro ni varovná proroctví, ekologické katastrofy, které implikují pocity nejistoty a obav. [Maříková, Petrusek, Vodáková a spol. 1996: 969–970]. Vzhledem k masové popularitě tohoto žánru vzniklo v myslích uživatelů mnoho sociálních reprezentací, jejichž původ pochází z vlastností sci-fi. Projevem takových sociální reprezentací jsou obavy z technologie budoucnosti, přehnaná očekávání od technologie, uvažování nad technikou, jako o potenciálně myslícím a emocionálním aktérovi. Tyto reprezentace mohou mít na uživatelovu důvěru pozitivní i negativní vliv. Tyto reprezentace jsou však zavádějící a nekorespondují se současnou situací. [Boyd, Holton 2017:].

Politické faktory

Uživatelé si umělou inteligenci často spojovali s nedostatečně nastaveným regulatorním rámcem. Nastavení právních systémů tak není připraveno na masovou integraci moderních technologií do každodenního života. Respondenti často jako příklad uváděli právní potíže se spuštěním samořídících vozidel do běžné provozu, nebo také často směřovali k otázkám právní odpovědnosti za případné škody. Další silný politický faktor byla možnost monitoringu – ať už státem, či jinou třetí osobou. Uživatelé jsou si tohoto rizika vědomi, avšak přístupy se liší. Někteří uživatelé se snaží blokovat možnosti sledování a prohlubují zabezpečení svých systémů. Jiní jsou rezignovaní, neboť se zastávají názor, že v dnešní době už není možné monitorování

Důvěra uživatelův umělou inteligenci, potažmo virtuální asistenty, je silně ovlivňována sociálními reprezentacemi. Do procesu jejich ukládání v uživatelově mysli vstupuje mnoho různých faktorů, které ovlivňují vnímání uživatele. Faktory, které jsem zde uvedl, jsou důležité aspekty, které uživatel vnímá při přímé, či nepřímé interakci s umělou inteligencí. Jejich základě dochází k tvorbě sociálních reprezentací a určování, zda reprezentace bude mít pozitivní, či negativní konotaci s daným subjektem zájmu. Výše zmíněné faktory jako: osobní zkušenosti, sci-fi, schopnosti umělé inteligence, regulatorní rámce – to vše jsou faktory, které člověk během svého života vnímá, a které ovlivňují jeho vnímání reality prostřednictví tvorby sociálních reprezentací. [Elcheroth, Doise, Reicher 2011: 730–742]

Závěr

Ve své práci jsem si vytyčil čtyři hlavní cíle. Prvním z cílů je určit hlavní zdroje důvěry v umělou inteligenci implementované do každodenního života. Na základě získaných dat jsem určil čtyři zdroje důvěry, které mohou být jak stálé, tak proměnlivé a které mají na uživatele výrazný vliv – *neutralita*, *víra v budoucnost*, *naplnění očekávání* a *blízkost*. Neutralitou je myšlen pocit neutrálního přístupu technologie k jedinci. Jedinec si neuvědomuje, že daná technologie byla stvořena člověkem za nějakým účelem, ale vidí pouze neutrální entitu, se kterou může komunikovat. Jedná se o stabilní zdroj důvěry, který má svůj původ ve společenské struktuře, tedy v kolektivním vnímání umělé inteligence určitou sociální skupinou. *Víra v budoucnost* také pochází ze společenské struktury. Někteří lidé mají tendenci k technologii přistupovat s optimismem, neboť jí vnímají jako něco nevyhnutelného s vysokým potenciálem být v mnoha ohledech lepší než člověk a pomáhat mu. Převaha takové vnímání ve společenské struktuře má za následek kolektivní důvěru, která se projevuje jako stálý zdroj důvěry u jednotlivců. Jedná se o slepou víru lidí v potenciál technologií, které jednoho dne budou téměř dokonalé a budou umět všechno. *Naplnění očekávání* je na rozdíl od předchozích dvou zdrojů důvěry zdroj proměnlivé. Je závislý na kumulované zkušenosti uživatele a v čase se značně proměňuje – důvěra se oslabuje, či prohlubuje na základě naplnění očekávání uživatele. Posledním zdrojem důvěry je *blízkost*. Blízkost má dvě podoby. První podobou je blízkost psychická, kdy uživatel dělá mnoho činností prostřednictvím virtuálního asistenta – pouští hudbu, televizi, volá, píše zprávy. Asistent dělá veškeré činnosti s ním a uživatel si k technologii vybuduje určitý vztah. Blízkost může mít podobu i fyzickou, neboť většina respondentů měla svého virtuálního asistenta neustále při ruce. Jsou jim neustálou oporou a to i v momentě, kdy telefon, či jiná technologii není přímo v jejich dosahu. Samotný fakt, že IVA uživatele konstantně propojuje s telefonem, může dát uživateli pocit, že jeho IVA je neustále s ním. To také podporuje Mirilelliho tezi, že si uživatelé mohou se svým mobilním telefonem vybudovat vztah, neboť s ním tráví hodně času. Stává se pro ně důležitým a hlídají si ho. [Mirilello a spol. 2013: 89–95] Na těchto základech si tak uživatel vytváří k technologii důvěru založenou na blízkosti. Jak už z podstaty zdroje vyplývá, i blízkost je podmíněná osobní zkušeností a je v čase proměnlivá, nelze ji tedy považovat za stálý zdroj důvěry.

Důvěra v umělou inteligenci je specifická a nelze k ní přistupovat stejně, jako například k důvěře mezi dvěma lidmi. Potvrdil jsem, že důvěra ve virtuálního asistenta může nabývat I. a II. dimenze dle Frederiksena, tedy že dimenze vztahů je založená na blízkosti vztahu na dostupnosti pomoci a vzájemné reciprocity (I. dimenze) a na předmětu důvěry – komu a v čem důvěřujeme (II. dimenze). [Frederiksen: 2012] Na druhou stranu III. dimenze, tedy identifikace s danou technologií zde není příliš možná. Ve vztahu k důvěře jsem také objevil silný vliv temporality. Po dobu, po kterou dochází

k naplňování očekávání, přičemž čím více má člověk s danou technologií zkušeností, tím více má s ní prožitků a tím spíše jí důvěřuje. Má důležitý vliv na proměnu důvěry u uživatelů, s níž se ale někteří autoři zabývající důvěrou v automatizaci vůbec nezaobírají [Rotter: 1967; Johns: 1996; Lee, See:2004; Liu, Yang, Xu: 2019].

Druhým cílem mé práce bylo zjistit, jak je důvěra v umělou inteligenci podmíněná sociálními reprezentacemi. Na základě získaných dat jsem potvrdil významný vliv sociálních reprezentací na myšlení jedinců a důvěru v umělou inteligenci. Na počátku jsem dokázal vytvořit tři kategorie představ uživatelů o virtuálních asistentech – *strojovost*, *personifikace* a *nehmatatelnost*. Na základě těchto představ jsem následně začal rozkrývat sociální reprezentace uživatelů o umělé inteligenci. Podařilo se mi dokázat, že důvěru skutečně reprezentace výrazně ovlivňují, a že je důvěra podmíněna sociálními reprezentacemi, které o vycházejí z kategorií *strojovost IVA, personifikace IVA a nehmatatelnost IVA*, a které tedy umělou inteligenci připodobňují k věcem, které uživatelé znají[41]. Mimo to jsem potvrdil, že sociální reprezentace jsou skutečně proměnné v čase (potvrzena dynamičnost sociálních reprezentací), kdy se důvěra prohlubovala na základě naplňování očekávání, či že docházelo k vzájemné výchově virtuálního asistenta a jeho uživatele. Také jsem potvrdil hierarchičnost sociálních reprezentací, kdy jednotlivé reprezentace nemají na důvěru jedince stejný vliv. Sociální reprezentace jádra mají větší vliv na jednotlivce, než sociální reprezentace periferie.

Dalším důležitým zjištěním je, že sociální reprezentace ukotvené v lidské mysli, mohou být v přímém rozporu – jedna sociální reprezentace vylučuje druhou, přičemž převaha jedné reprezentace nad druhou závisí na konkrétní situaci. Nebylo například výjimkou, že si uživatelé přáli, aby měl jejich virtuální asistent smysl pro humor, byl empatický a dokázal poradit. Na straně druhé však uváděli, že nechce, aby IVA měl emoce a aktivně se ptal a řešil osobní věci. Protichůdnost svých výroků následně nedokázal vysvětlit. Protichůdné reprezentace se projevují neuceleným, či protichůdným smýšlením o umělé inteligenci.[42] S variantou protichůdných sociálních reprezentací příliš nepracují ani autoři uvedení v teoretické části práce. [Moscovici: 1988; Höijer 2011: 7–12; Elcheroth, Doise, Reicher: 2011] Naopak Marková a spol. protichůdnost připouštějí, přičemž se projevuje nejasným a neuceleným pohledem na určitou skutečnost. [Marková a spol.: 1998]

Dalším cílem mé práce je zjistit, jaká rizika si jedinec uvědomuje ve vztahu k umělé inteligenci. V rámci mé práce jsem určil šest hlavních rizik, které si uživatelé uvědomují, jsou to: *nepředpokládaná chyba softwaru*, *mechanická chyba*, *zneužití IVA třetí stranou*, *právní rizika*, *strach*

[41] Například: Strojovost jako mobilní telefon, personifikace jako otrok, nehmatatelnost jako zdrojový kód.

[42] Například protichůdné smýšlení a odpovědi u respondentky Anety ve vztahu k IVA

z odcizení a *ovládnutí technologiemi. Nepředpokládanou chybou softwaru* se rozumí neočekávaná chyba, která nastane nesprávným nastavením softwaru, či interní chyba, která zapříčiní nesprávné fungování technologie. Zde jsem vyvrátil Madhavanovo a Wiegmannovo pojetí důvěry a rizik v technologie, kdy uživatelé jsou podle autorů mnohem vnímavější k chybám strojů než k chybám lidí [Madhavan, Wiegmann: 2007]. Dle mého výzkumu mají uživatelé tendenci chyby technologie spíše omlouvat. To potvrzuje i další riziko. *Mechanická chyba* je chyba očekávatelná. Postupné opotřebení materiálu, či čidel, zpomalování fungování, které vyústí v nesprávné fungování IVA. Riziko, které si také uvědomuje většina uživatelů a které většina uživatelů omlouvá. Co je pro uživatele důležité, je *riziko zneužití třetí stranou*. Zde potvrzuji velkou citlivost uživatelů ke ztrátě osobních údajů. Na druhou stranu, uživatelé se nemohou shodnout, do jaké míry za to nese zodpovědnost IVA. Pro řadu uživatelů je to jen snadno zneužitelný nástroj, ale jeho zneužití třetí stranou ještě nepovažují všichni uživatelé za důvod pro celkovou nedůvěru k virtuálním asistentům. Uživatelé si také uvědomují nedostatečné nastavení systému, které má za následek *právní rizika*. Někteří uživatelé projevili strach z *rizika odcizení*. Tedy, že lidé budou technologie využívat více, na úkor mezilidské interakce. Jako poslední riziko se často objevovala možnost *ovládnutí technologiemi*. Jedná se o strach uživatelů z přílišných schopností umělé inteligence, které může vést k její vzpouře proti lidem a ohrožení lidského druhu jako takového. Přestože si respondenti uvědomují, že umělá inteligence ještě není na takové úrovni, tento strach je pro ně reálný a aktivně zasahuje do jejich interakce s IVA (nesvěřují mu určité úkoly, či ho vůbec nepropojují s nějakou jinou technologií, nebo vůbec nechtějí další rozvoj umělé inteligence po emoční stránce). Rozličné vnímání rizik vychází ze sociálních reprezentací, které mají svůj původ v sociální, ekonomických, kulturních, či politických faktorech. Všichni respondenti jsou aktivními uživateli virtuálních asistentů, což znamená, že výzkumný vzorek je o jisté míry homogenní a ukazuje pouze zdroje důvěry a rizika relevantní pouze pro tuto skupinu.

Práce přináší mnoho témat a zajímavých postřehů, které mohou být námětem pro další zkoumání. Jedním z témat je třeba zjistit prostřednictví sociálních reprezentací, jak uživatelé IVA vidí okolní svět. Tedy v čem je jejich vnímání světa odlišný od jiných sociálních skupin. Ve své práci docházím k 3 kategoriím IVA – *strojovost*, *personifikace*, *nehmatatelnost*, které jsou určitým odrazem jejich vnímání jedinci. Bylo by zajímavé zkoumat, zda tyto kategorie jsou platné i v jiné sociální skupině, například lidí, co IVA využívat odmítají. Nebo je možné zaměřit pozornost přímo na sociální reprezentace o umělé inteligenci v rámci sociálních struktur a následně je mezi sebou porovnávat. Dále mohou být má zjištění o vnímání IVA některými uživateli využity i na poli Science and Technology Studies. Zejména informace popisující, jak uživatelé hovoří o vzájemném vztahu s virtuálním asistentem, vzájemnému vychovávání a vzájemnému porozumění. Podobná zjištění mohou výrazně přispět k pochopení vzájemného vztahu člověka a umělé inteligence. V návaznosti

přímo na tuto práci je možné další rozdělení uživatelů virtuálních asistentů, například podle jejich vztahu k IVA. Podle výsledků mé práce v této sociální skupině převládá určitý technooptimismus. Přesto však vnímání IVA nemusí být stejné. Respektive by bylo zajímavé zjistit, v jakých ohledech jsou uživatelé IVA spíše technooptimističtí a v čem spíše technopesimističtější ve vztahu k umělé inteligenci.

Summary

In my thesis, I set four main goals. The first goal is to identify the main sources of trust in the artificial intelligence implemented in everyday life. Based on the data obtained, I identified four main sources of trust that have a significant impact on users – *neutrality, belief in the future, fulfillment of expectations* and *closeness*. By *neutrality* is meant the feeling of technology's neutral approach to the individual. It is a stable source of trust, which has its origin in the social structure, ie in the collective perception of artificial intelligence by a certain society. *Belief in the future* also comes from the social structure. Some people tend to approach technology with optimism because they perceive it as something inevitable with high potential and potential for application. The predominance of such perceptions in the social structure results in collective trust, which manifests itself as a constant source of trust in individuals. It is a blind belief of people (confidence) in the potential of technologies that will one day be almost perfect and will know everything. *Fulfillment of expectations*, unlike the previous two sources of trust, is a volatile source. It depends on the accumulated experience of the user and changes considerably over time - trust is weakened or deepened by meeting the user's expectations. The last source of trust is *closeness*. Closeness has two forms. The first form is mental closeness, where the user does many activities through a virtual assistant - he plays music, watches television, calls, and writes news. The assistant does all the work with it and the user builds a relationship with the technology. Closeness can also be physical, as most respondents always had their virtual assistant at hand. They are a constant support to them, even when the phone or other technology is not directly within their reach. The fact that the IVA constantly connects the user to the phone can give the user the feeling that his IVA is constantly with him. This also supports Mirilelli's thesis that users can build a relationship with their mobile phone because they spend a lot of time with it. It becomes important for them and they watch over it. [Mirilello a spol. 2013: 89–95] On this basis, the user builds a trust based on closeness to the technology. As follows from the nature of the source, closeness is also conditioned by personal experience and is variable over time, so it cannot be considered a permanent source of trust.

Trust in artificial intelligence is specific and cannot be approached in the same way as, for example, trust between two people. For example, I have confirmed that trust in a virtual assistant can take on the 1st and 2nd dimensions according to Frederiksen that is, the dimension of relationships is based on the proximity of the relationship to the availability of help and mutual reciprocity (1st dimension) and on the object of trust - to whom and in what we trust (2nd dimension). [Frederiksen: 2012] On the other hand 3rd dimension, ie identification with the given technology is not very possible here. In relation to trust, I also discovered a strong influence of temporality, , which has an important effect on the change of trust in users, but with which some authors dealing with trust in automation do not

work at all [Rotter: 1967; Johns: 1996; Lee, See:2004; Liu, Yang, Xu: 2019]. On the contrary, Marková et al. they admit contradiction, manifesting itself in a vague and incoherent view of a certain fact. [Marková a spol.: 1998]

The second goal of my work was to find out whether trust in artificial intelligence is conditioned by social representations. Based on the obtained data, I confirmed the significant influence of social representations on the thinking of individuals and confidence in artificial intelligence. At the beginning, I created three categories of user ideas about virtual assistants – *mechanicality, personification* and *intangibility*. Based on these ideas, I subsequently began to reveal the social representations of users about artificial intelligence. I managed to prove that trust is really significantly influenced by representations, and that trust is conditioned by social representations that are based on categories *mechanicality IVA, personification IVA* and *intangibility IVA*. They liken artificial intelligence to things that users know[43]. In addition, I confirmed that social representations are indeed variable over time (confirmed dynamism of social representations) when the trust deepened on the basis of fulfilling expectations, or that there was a mutual education of the virtual assistant and his user. I also confirmed the hierarchical nature of social representations, where individual representations do not have the same effect on an individual's trust. The social representations of the core based on the machine structure of the IVA give a certain picture of a virtual assistant. Thinking about its mechanical and software disadvantages are the social representations of the periphery that do not affect trust so much. Opposing representations are manifested by incoherent or contradictory thinking about artificial intelligence.

Another important finding is that social representations rooted in the human mind can be in direct conflict – one social representation excludes the other, with the predominance of one representation over the other depending on the specific situation. In my research, I have found that such a situation can occur and the user may not be aware of it. The user wanted his virtual assistant to have a sense of humor, to be empathetic and to be able to advice. After a few sentences, however, he stated that he did not want the IVA to have emotions and to actively ask questions and deal with personal matters. User could not subsequently explain the contradiction of his statements.[44] The authors mentioned in the theoretical part of the thesis do not work too much with the variant of opposite social representations [Moscovici: 1988; Höijer 2011: 7–12; Elcheroth, Doise, Reicher: 2011].

Another goal of my work is to find out what risks the individual is aware of in relation to artificial

[43] For example: Machine as a mobile phone, personification as a slave, intangibility as a source code.

[44] For example, the contradictory thinking and answers of respondent Aneta in relation to IVA

intelligence. In my work, I have identified six main risks that users are aware of are: *unexpected software error*, *mechanical error*, *misuse of IVA by a third party*, *legal risks*, *fear of alienate* and *control by technology*. *An unexpected software error* is an unexpected error that occurs due to incorrect software settings or an internal error that causes the technology to malfunction. Here I refuted Madhavan's and Wiegmann's notions of trust and risk in technology, where users say users are much more receptive to machine errors than to human errors. [Madhavan, Wiegmann: 2007]. According to my research, users tend to apologize of technology errors. This confirms another risk. A *mechanical error* is an expected error. Gradual wear of material or sensors, slowing down of operation, which results in incorrect functioning of IVA. A risk that most users are also aware of and which most users apologize. What is important for users is the risk of *misuse by a third party*. Here I confirm the great sensitivity of users to the loss of personal data. On the other hand, users cannot agree on the extent to which the IVA is responsible for this. For many users, it is only an easily exploited tool, but its misuse by a third party is not yet considered by all users as a reason for a total distrust of virtual assistants. Users are also aware of insufficient system settings, which results in *legal risks*. Some users have expressed fears of the *risk of theft*. That is, people will use technology more, at the expense of interpersonal interaction. The last risk was often the possibility of *control by technology*. It is the fear of users of excessive artificial intelligence capabilities, which can lead to its rebellion against humans and endanger the human species as such. Although the respondents are aware that artificial intelligence is not yet at such a level, this fear is real for them and actively interferes with their interaction with IVA (they do not entrust certain tasks to it, or connect it with any other technology, or do not want further emotionally development of artificial intelligence at all). Different perceptions of risks are based on social representations, which have their origin in social, economic, cultural or political factors. All respondents are active users of virtual assistants, which means that the research sample is somewhat homogeneous and shows only sources of trust and risks relevant only to this group.

The work brings many topics and interesting observations that can be a topic for further research. One of the topics is to find out, through social representations, how IVA users see the world around them. So how is their perception of the world different from other social groups. In my work I come to 3 categories of IVA - *machine*, *personification*, *intangibility*, which are a certain reflection of their perception by individuals. It would be interesting to examine whether these categories are valid in another social group, such as people who refuse to use IVA. Or it is possible to focus attention directly on social representations about artificial intelligence within social structures and then compare them with each other. Furthermore, my findings on the perception of IVA by some users can be used in the field of Science and Technology Studies. In particular, information describing how users talk about the relationship with the virtual assistant, mutual upbringing, and mutual

understanding. Similar findings can significantly contribute to understanding the relationship between man and artificial intelligence. Following this work, it is possible to further divide the users of virtual assistants, for example according to their relationship to IVA. According to the results of my work, a certain techno-optimism prevails in this social group. Nevertheless, the perception of IVA may not be the same. Respectively, it would be interesting to find out in which respects IVA users are more techno-optimistic and in what more techno-pessimistic in relation to artificial intelligence.

Použitá literatura

Abric, J. C. 1996. *Specific processes of social representations.* Université de Provence, Aix-en-Provence, Francie.

Aluwihare-samaranayake, D. 2012. „Ethics in Qualitative Research: A View of the Participants' and Researches' World from a Critical Standpoint." *Journal of Qualitative Methods,* 11 (2).

Bartholomew, K., Henderson, A. J.Z., Marcia, J.E. 2000. *Coding semi-structured interviews in social psychological research.* Cambridge: Cambridge University Press

Beck, U. 2018. *Riziková společnost. Na cestě k jiné modernitě*. SLON: Praha

Beer, D., Burrows, R. 2007. *Sociology and, of and in Web 2.0: Some Initial Considerations*. Sociological Research Online, 12(5).

Belanger, F., Xu, H. 2015. „The role of information systems research in shaping the future of information privacy." *Information systems journal*, 25(6).

Blascovich, J. 2002. „Social Influence within Immersive Virtual Environments." *Computer Supported Cooperative Work.* CSCW

Boyd, D., Crawford, K. 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15 (5).

BOSTRÖM, N. 2017. *Superinteligence*. Praha: Prostor

Brill, T. M., Munoz, L., Miller, R. J. 2019. Siri, Alexa, and other digital assistants: a study of customer satisfaction with artificial intelligence applications. *Journal of Marketing Management*, 35(15-16)

Elcheroth, G., Doise, W., Reicher, S. 2011. „On the Knowledge of Politics and the Politics of Knowledge: How a Social Representations Approach Helps Us Rethink the Subject of Political Psychology." *Political Psychology*, 32 (5).

Eriksson, E. H. 2002. Dětství a společnost. Praha: Argo

Frederiksen, M. 2012. „Dimensions of trust: An empirical revisit to Simmel's formal sociology of intersubjective trust." *Current Sociology*, 60(6)

French, R. M. 1990. „Subcognition and limits of the Turing Test." *Mind,* 99 (393).

Galletta, A. 2013. *Mastering the Semi-Structured Interview and Beyond*. New York university press: New York & London.

Giddens, A. 2003. Důsledky modernity. Praha: SLON

Harper, D., Thompson, A. R. 2011. *Qualitative Research Methods in Mental Health and Psychotherapy.* Oxford: John Wiley & Sons.

Hendl, J., Remr, J. 2017. Metody výzkumu a evaluace. Praha: Portál.

Hengstler, M., Enkel, E., Duelli, S. 2016. „Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices." *Technological Forecasting & Social Changez,* 105 (1).

Hoff, K. A., Bashir, M. 2014. Trust in Automation. *Human Factors: The Journal Of the Human Factors and Ergonomics Society,* 57 (4).

Höijer, B. 2011. Social Representations Theory. A New Theory for Media Research. *Nordicom Review,* 32 (2).

Hollway, W., Jefferson, T. 2008. „Free association narrative interview. Given LM (ed.)" *The Sage Encyclopedia of Qualitative Research Methods.* London: SAGE.

Chung, H., Iorga, M., Voas, J., Lee, S. 2017. „Axela, Can I Trust You?“ *Computer*, 50 (9).

Johns, J. L. 1996. A concept analysis of trust. *Journal of Advanced Nursing,* 24 (1).

Kim, K., Boelling, L., Haesler, S., Bailenson, J., Bruder, G., Welch, G. F. 2018. *Does a Digital Assistant Need a Body? The Influence of Visual Embodiment and Social Behavior on the Perception of Intelligent Virtual Agents in AR 2018*. IEEE International Symposium on Mixed and Augmented Reality (ISMAR).

Kyslova, O., Bernyk, E. 2014. „New Media as a Formation Factor for Digital Sociology: The Consequences of the Networking in the Society and the Intellectualization of the Communications.“ *Studies of Changing Societies*, 2013(3).

Lash, S. 2006. Dialectic of information? A response to Taylor. *Information, Communication & Society*, 9(5).

Lee, J. D., See, K. A. 2004. „Trust in automation: Designing for appropriate reliance.“ *Human Factors: The Journal of the Human Factors and Ergonomics Sociely*, 46 (1).

Liu, P., Yang, R., & Xu, Z. (2019). Public acceptance of fully automated driving: Effects of social trust and risk/benefit perceptions. *Risk Analysis*, 39 (2).

Lloyd, B., Dunveen, G.1992. *Gender identities and education: The impact of starting school*. London: Harvester-Wheatsheaf.

Luhmann, N. 1979. *Trust and power*. Avon: Pitman press.

Lupton, D. 2014a. *Self-tracking Cultures: Towards a Sociology of Personal Informatics*. Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures the Future of Design - OzCHI ’14.

Lupton D. 2014b. „Self-Tracking Modes: Reflexive Self-Monitoring and Data Practices.“ *SSRN Electronic Journal*.

Madhavan, P., Wiegmann, D. A. 2007. „Similarities and differences between human–human and human–automation trust: an integrative review.“ *Theoretical Issues in Ergonomics Science*, 8 (4).

Marková, I. A spol. 1998. „Social representations of the individual: a post-Communist perspective.“ *European Journal of Social Psychology*, 28 (5).

Maříková, H., Petrusek, M., Vodáková, A. 1996. *Velký sociologický slovník*. Karolinum: Praha

McIntosh, J., Morse, J. M. 2015. *Situating and Constructing Diversity in Semi-Structured Interviews.* Research Article.

Merrilees, B., Fry, M.-L. 2003. „E-trust: the influence of perceived interactivity on e-retailing users“ *Marketing Intelligence & Planning*, 21 (2).

Miritello, G. a spol. 2013. „Time as a limited resource: Communication strategy in mobile phone networks.” *Social Networks,* 35 (1).

Moscovici, S. 1988. Notes towards a description of Social Representations. *European Journal of Social Psychology*, 18 (3).

Morse, J. M., Field, P. A. 1995. *Qualitative research methods for health professionals*. Psychology.

Nass, C., a spol. 1995. *Can Computer Personalities Be Human Personalities?* Stanford University.

O’Connor, C. 2012. *Using Social representations theory to examine lay explanation of contemporary social crises: The case of Ireland’s recession*. University College: London.

Orb, A., Eisenhauer, L., Wynaden, D. 2001. „Ethics in qualitative research*.” Journal of Nursing Scholarship*, 33 (1).

Peggs, K. 2013. The "Animal-Advocacy Agenda": Exploring Sociology for Non-Human Animals. *The Sociological Review*, 61 (3).

Purwanto, P., Kuswandi, K., & Fatmah, F. (2020). Interactive Applications with Artificial Intelligence: The Role of Trust among Digital Assistant Users. *Foresight and STI Governance*, 14 (2).

Rezaev, A., Tregubova, N. 2018. *Are sociologists ready for artificial sociality? Current issues and future prospects for studing artificial intelligence in the social sciences*. Monitoring of Public Opinion.

Rotter, J. B. 1967. „A new scale for the measurement of interpersonal trust." *Journal of Personality* 35 (4).

Rousseau, D. M., Sitkin, S.B., Burt, R. S., Camerer, C 1998. „Not So Different After All: A Cross-Discipline View Of Trust." *Academy of Management Review*, 23 (3).

Russell, S., Norvig, P. 2010. *Intelligence artificielle: Avec plus de 500 exercices*. PEARSON.

Sedláček, J. 1979. *Východiska Durkheimovy sociologie*. Praha: Univerzita Karlova.

Shamoo, A. E., Resnik, D. B. 2009. *Responsible conduct of research*. New York: Oxford University Press.

Shapiro, S. C. 1992. *Encyclopedia of artificial intelligence second edition. Volume 1.* A Wiley-Interscience Publication: New York.

Simmel, G. 1997. *Peníze v moderní kultuře a jiné eseje*. Praha: SLON

Smith, N., Joffe, H. 2012. *How the public engages with global warming: A social representations approach*. Public Understanding of Science 22 (1).

Strauss, A. L., Corbin, J. 1999. *Základy kvalitativního výzkumu: postupy a techniky metody zakotvené teorie*. Brno: Sdružení Podané ruce.

Sun, T. Q., Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, 36 (2).

Wanger, W., Valencia, J. 1996. „Relevance, discourse and the 'hot' stable core social representations — A structural analysis of word associations." *British Journal of Social Psychology*, 35 (3).

Wanger, W. a spol. 1999. „Theory and Method of Social Representations." *Asian Journal of Social Psychology* 2 (1).

Internetové zdroje

„Apple.com/siri". 2020.apple.com . [on-line] [cit. 19.3. 2020]. Dostupné z: https://www.apple.com/siri/

„Applenovinky". 2019. applenovinky.cz. [on-line] [cit. 8.7. 2020].Dostupné z: https://applenovinky.cz/2019/04/vyzkousejte-siri-v-cestine/

„*Česko-Slovenská filmová databáze*". 2001. CSFD. [on-line] [cit. 15.7. 2020]. Dostupné z: https://www.csfd.cz/film/312650-ona/prehled/

„*IBM – Artificial Intelligence*.". Tisková zpráva. [on-line] Praha: 15. října 2018.

„Orientace v GDPR" 2020. mvcr.cz. [on-line] [cit. 8.7. 2020]. Dostupné z: https://www.mvcr.cz/gdpr/clanek/zvlastni-kategorie-osobnich-udaju.aspx

„*Preliminary report – Highway HWY18MH010*" 2018. The Nexus. University of Oregon. [on-line] [cit. 10.7. 2020]. Dostupné z: https://www.urbanismnext.org/resources/preliminary-report-highway-hwy18mh010

„*The Third-Generation Web is Coming*." 2006. In Kurzweil accelerating intelligence. [on-line] [cit. 14.3. 2020]. Dostupné z: https://www.kurzweilai.net/the-third-generation-web-is-coming

Teze Diplomové práce

TEZE DIPLOMOVÉ PRÁCE

Univerzita Karlova v Praze

Fakulta sociálních věd
Institut sociologických studií
Katedra sociologie

Sociologické zkoumání zdrojů důvěry v umělou inteligenci

Student: Jakub Janouš

Konzultant: Dino Numerato

Ve své diplomové práci se budu zabývat zdroji důvěry v umělou inteligenci, která je různým způsobem implementovaná do každodenního života. Vzhledem k stále se zvyšujícímu významu umělé inteligence v různých odvětví lidské činnosti je toto téma nanejvýš aktuální a relevantní.

Moje práce bude vycházet z průzkumu veřejného mínění, který probíhal na konci roku 2018, a který si objednala společnost IBM u výzkumné agentury MNS Market Research. Z výzkumné zprávy je patrné, že osobní zkušenost s umělou inteligencí má v České republice pouhých 19 % respondentů, což je nejméně ze 4 celkově zkoumaných států[45]. Naopak důvěru v umělou inteligenci potvrdilo více než polovina dotazovaných, konkrétně 56 % respondentů. Je tedy otázkou v čem spočívá ona důvěra v umělou inteligenci, kterou má 56 % dotazovaných, přesto že osobní zkušenost s ní uvedlo pouhých 19 % dotazovaných. Z tohoto poznatku také vyplývá další zajímavá informace. Podle stejného průzkumu 67 % dotazovaných v České republice uvedlo

[45] Výzkum zahrnuje tyto státy: Česká republika, Maďarsko, Polsko a Rusko.

pozitivní přístup k inovacím v umělé inteligence. Tento fakt opět plyne z důvěry zkoumaného celku v umělou inteligenci. Naopak nízká osobní zkušenost s umělou inteligencí spíše naznačuje neznalost respondentů s umělou inteligencí, neboť v současné době je umělá inteligence běžnou součástí například internetového vyhledávače, či androidu v mobilním zařízení.

Důvěru v umělou inteligenci budu zkoumat optikou teorie sociálních reprezentací. V návaznosti na Moscoviciho (1984) nahlížím na sociální reprezentace jako na soubor organizovaných a strukturovaných poznatků o specifickém sociálním fenoménu, který činí pro člověka sociální realitu srozumitelnou, bezpečnou a možnou ji sdílet v procesu komunikace s ostatními jedinci. Tento teoretický koncept by mohl[46] vysvětlovat, nebo alespoň přiblížit důvody důvěry v něco, s čím přišlo do vědomého styku pouhých 19 % respondentů.

Výzkumná otázka

Tato práce si klade za cíl zkoumat a určit hlavní zdroje důvěry v umělou inteligenci, a to pomocí těchto výzkumných otázek: Jaké jsou hlavní zdroje důvěry v umělou inteligenci implementované do každodenního života? Jaké jsou sociální reprezentace umělé inteligence? Jaká rizika si jedinec uvědomuje ve vztahu k umělé inteligenci? Měla by se umělá inteligence dále rozvíjet? Do jaké míry je důvěra v umělou inteligenci ovlivněna sociálními, politickými, ekonomickými a kulturními faktory?

Metodologický postup

Pro pochopení zdrojů důvěry či nedůvěry v umělou inteligencí využiji kvalitativní metodu zkoumání, která umožňuje hlubší vhled do určité situace či problematiky za účelem zjištění její podstaty – v tomto ohledu podstaty důvěry v umělou inteligenci. Prostřednictvím polo-strukturovaných výzkumných rozhovorů budu zjišťovat názory a pohledy na umělou inteligenci. Rozhovory budou v případě potřeby obohaceny o pozorování. Následně budu pomocí definovaných kódů hledat určité znaky, které budou charakterizovat důvěru, či naopak nedůvěru jedinců v umělou inteligenci.

Předběžná struktura práce:

1. Obsah
2. Úvod
3. Teoretická část
    3.1. Proces racionalizace ve 20. a 21. století.
    3.2. Definice umělá inteligence a její rozšíření v každodenním životě
    3.3. Teorie sociální reprezentace
    3.4. Teorie důvěry v podání různých autorů
    3.5. Výsledky průzkumu veřejného mínění od společnosti MNS Market research
4. Metodologie
    4.1. Cíle a výzkumné otázky
    4.2. Výzkumný postup
    4.3. Výběr participantů

[46] V tuto chvíli nelze říci, nakolik je tato teorie správná.

4.4. Charakteristika rozhovorů

5. Analytická část
6. Závěr
7. Literatura

Předběžný seznam literatury:

BOSTROM, N. 2017. Superinteligence. Praha: Prostor
BRYNJOLFSSON, E. A MCAFEE, A. 2015. Druhý věk strojů: práce, pokrok a prosperita v éře špičkových technologií. Brno: Jan Melvil Publishing.
GIDDENS, A. 1998. Důsledky modernity. Praha: Sociologické nakladatelství.
HENDL, J. 2005. Kvalitativní výzkum: základní metody a aplikace. Praha: Portál
HENDL, J. a REMR, J. 2017. Metody výzkumu a evaluace. Praha: Portál
LUHMANN, N. 1979. Trust and power. Avon: Pitman Press
LUHMANN, N. 2006. Sociální systémy. Nárys obecné teorie. Centrum pro studium demokracie a kultury.
MAŘÍKOVÁ, H. PETRUSEK, M. a VODÁKOVÁ, A. 1996.Velký sociologický slovník. Praha: Karolinom.
MLYNÁŘ, J., ALAVY, H. S., VERMA, H., & CANTONI, L. 2018. Towards a Sociological Conception of Artificial Intelligence. In *International Conference on Artificial General Intelligence* (pp. 130-139). Springer, Springer Nature Switzerland.
MOSCOVICI, S. 2000. Social Representations. Explorations in Social Psychology. Cambridge: Polity Press.
SZTOMPKA, P, P. (2006). New perspectives on trust. *American Journal of Sociology*, *112*(3): stránky???
SZTOMPKA, P. 1999. Trust. A Sociological Theory. Cambridge: Cambridge University Press.
SZTOMPKA, P. (2007). Trust in science: Robert K. Merton's inspirations. *Journal of Classical Sociology*, detaily k článku???

WEBER, M. 2009a. „Předznamenání k sebraným statím k sociologii náboženství“ In Metodologie, sociologie a politika, Praha: Oikoymenh
WEBER, M. 2009b. „Protestantská etika a duch kapitalismu“ In Metodologie, sociologie a politika, Praha: Oikoymenh.

Podpis konzultanta: _____

Podpis studenta: _____

Seznam příloh

Příloha č. 1: Výzkumný online filtrovací dotazník

Výzkumný dotazník – využívání inteligentních virtuálních asistentů

Tento krátký dotazník slouží k výzkumu přístupu a využívání inteligentních virtuálních asistentů (IVA) lidmi. Na základě odpovědí v tomto dotazníku budou někteří respondenti vybráni a požádáni o účast v osobních výzkumných rozhovorech pro upřesnění či doplnění odpovědí v tomto dotazníku.
Výstupy z tohoto dotazníku budou zpracovány, analyzovány a následně zveřejněny v mé diplomové práci. Získaná data nebudou sloužit k žádným jiným než akademickým účelům, ani nebudou poskytována třetím stranám a před jejich prezentací budou participanti výzkumu anonymizováni.
Vyplněním a odesláním dotazníku dotazovaný souhlasí se zpracováním dat, tak jak bylo popsáno výše, a souhlasí s jejich následným využitím v mé diplomové práci.

Inteligentní virtuální asistent (IVA) je aplikace či program využívající algoritmy umělé inteligence, který je uživateli oporou v konkrétních úkonech. (Jedna z možných definic IVA podle Luca Lamontagne) Jedná se o aplikaci, kterou můžete mít nastavenou ve Vašem mobilním zařízení, nebo nainstalovanou v počítači.

1. Už jste někdy slyšel/a pojem inteligentní virtuální asistent?
   - ANO
   - NE
   - NEVÍM

2. Které z těchto platforem si myslíte, že poskytují možnost využívání IVA? (může být zaškrtnuto více možností)
   - Amazon
   - Apple
   - Facebook
   - Google
   - Microsoft
   - Samsung
   - Yandex
   - Jiné?

3. Znáte některé z těchto virtuálních asistentů? (může být zaškrtnuto více možností) Pokud neznáte žádného, přeskočte na otázku 6
   - Alisa
   - Alexa
   - Bixby
   - Google Assistant
   - M
   - Cortana
   - Siri
   - WeChat

4. Pracoval/a jste s některým z výše uvedených virtuálních asistentů? Se kterým? (Napsat)

5. Používáte inteligentního virtuálního asistenta?
   - Nepoužívám
   - Jednou, nebo párkrát jsem to zkusil/a
   - Alespoň jednou za měsíc ho využiji

- Používám ho alespoň jednou za týden
- Používám ho každý den

6. Co si myslíte o následujících výrocích. 1 – ANO, 2 – Nemám názor, 3 – NE.

   IVA je užitečný nástroj pro uživatele. 1 2 3
   IVA je zbytečný nástroj pro uživatele. 1 2 3
   Mohl by uživatel prostřednictvím IVA řídit domácnost i v jeho nepřítomnosti? 1 2 3
   Dovolil/a byste IVA, aby sám objednal jídlo, které dochází v ledničce? 1 2 3
   Nechal/a byste IVA vypnout televizi v předem definovaný čas? (např.: v 01:00) 1 2 3
   Nechal/a byste si od IVA vyhledat nejrychlejší cestu do práce? 1 2 3
   Nechal/a byste IVA, aby hlídal, kdy má kdo narozeniny? 1 2 3
   Nechal/a byste IVA, aby půl hodiny před Vaším příchodem z práce předehřál troubu? 1 2 3
   Nechal/a byste IVA, aby Vám spravoval kalendář? 1 2 3
   Nechal/a byste IVA řídit autonomně řízené vozidlo? 1 2 3

7. Důvěřujete Inteligentní virtuálním asistentům?
   - Nedůvěřuji
   - Spíše nedůvěřuji
   - Ani nedůvěřuji ani důvěřuji
   - Spíše důvěřuji
   - Důvěřuji

8. Vaše pohlaví?
   - Muž
   - Žena
   - Transgender

9. Váš věk
   - 1522
   - 23–37
   - 38–52
   - 53–74
   - 75 +

10. Byl/a byste ochotný/á zúčastnit se osobního výzkumného rozhovoru na toto téma?
    - ANO
    - NE

V případě zájmu o osobní výzkumný rozhovor, prosím o uvedení kontaktu, na kterém Vás mohu kontaktovat.

Příloha č. 2: Scénář k výzkumným rozhovorům

Scénář k výzkumným rozhovorům

I část

- Proč využíváš svého virtuálního asistenta?
    - Co od něj očekáváš?
    - Co tě přimělo ho začít využívat?
    - Co všechno na něm děláš?

- Jaká je tvoje první asociace, když se řekne inteligentní virtuální asistent?

- Jaké jsou výhody IVA využívat?
- Jaké jsou nevýhody IVA využívat?
- Jsou schopnosti IVA v současnosti dostatečné?
    - Co bys chtěl, aby IVA všechno umělo? (*vlastnosti*)
- Zklamal tě někdy tvůj IVA?
    - Změnil to tvůj pohled na něj?
    - Co by se muselo stát, aby si mu přestal důvěřovat?

II. část

- Jak by si popsala vztah mezi tebou a IVA?

- Používá někdo z tvého okruhu lidí také IVA?

    - Proč?
- Jak si myslíš, že jsou IVA mezi uživateli vnímáni?
    - Staří
    - Mladí
    - Kamarádi
    - Odborníci

- Co byl nejsilnější okamžik, který se ti s IVA stal?
    - Co dalšího se ti stalo?
    - Pozitivní/negativní?
    - Slyšel si o nějakých incidentech při využívání IVA?

- Jsou nějaké návyky /chování, které si změnil kvůli využívání IVA?

- Jak vnímáš, když někoho vidíš, že využívá IVA?
    - Je ti to
    - Potřebuješ něco zjistit, využila by si služeb IVA?

- Co podle tebe nejvíce ovlivňuje tvůj pohled na IVA

III. část

- Může být využívání IVA nebezpečné?
    - Jak?
    - Proč?
- Existují nějaká rizika?

IV. část

- Měl by se IVA ještě více implementovat do každodenního života?
    - Kde by bylo vhodné využití?
- Jak ideálně by si ho chtěl využívat?
- Kde by neměl být IVA implementován?
    - Proč?
- Co si myslíš o řízení domácnosti přes telefon?
    - Světla, trouba, televize, topení, časovače.
- V dotazníku: ANO: narozeniny, kalendář, časovače, troubu, cestu do práce
    - NE: autonomní vozidla

Příloha č. 3: Informovaný souhlas

Informovaný souhlas

Tento rozhovor slouží k výzkumu vnímání umělé inteligence. Na základě dotazníkového filtračního šetření, jste byl/a vybrán jako vhodný respondent do výzkumu zabývajícího se zkoumáním přístupu lidí k umělé inteligenci, která je implementovaná do každodenního života.

Výstupy z tohoto výzkumného rozhovoru budou zpracovány, analyzovány a následně zveřejněny v mé diplomové práci. Získaná data nebudou sloužit k žádným jiným než akademickým účelům, ani nebudou poskytována třetím stranám a před jejich prezentací budou participanti výzkumu anonymizováni.

Souhlasíte s tímto výzkumným rozhovorem, s pořízením audio záznamu, se zpracováním a analyzováním získaných dat, s jejich anonymizací pomocí změny Vašeho jména a následným zveřejněním v mé diplomové práci, která bude veřejně dostupná v digitálním repositáři závěrečných prací Univerzity Karlovy.

Souhlasíte se všemi výše zmíněnými body, tak jak jsou výše popsány? Pokud souhlasíte, prosím podepište tento informovaný souhlas.

_____