

UNIVERZITA KARLOVA

KATOLICKÁ TEOLOGICKÁ FAKULTA

Katedra systematické a pastorální teologie

Ing. Kateřina Švarcová

# **ETIKA A UMĚLÁ INTELIGENCE**

Diplomová práce

Vedoucí práce: Mgr. Ing. Prokop Sousedík, Ph.D.

Praha 2024



## **Prohlášení**

1. Prohlašuji, že jsem předkládanou práci zpracovala samostatně a použila 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 2.5.2024

Kateřina Švarcová

## **Bibliografická citace**

Etika a umělá inteligence [rukopis]: diplomová práce / Kateřina Švarcová; vedoucí práce: Prokop Sousedík. --Praha, 2024. -- 66 s.

## **Anotace**

Diplomová práce se zabývá problematikou etiky umělé inteligence (dále také „UI“). Práce je rozdělena do tří základních částí. V první části jsou popsány základní pojmy. Zejména jsou zde vysvětleny základní principy algoritmů a fungování UI po technické stránce. V další části se práce pomocí textů J.R. Searleho a J. Peregrina zabývá srovnáním UI a lidského myšlení z filozofického pohledu. Z této části je vyvozen závěr, že UI je stroj a jako takový nemá morální zodpovědnost. Morální zodpovědnost při práci s UI mají lidé, a to vývojáři UI a její uživatelé. Tímto se práce zabývá ve třetí části, která představuje různé aktuální morální problémy a na základě utilitaristického a deontologického směru stanovuje základní etické principy pro práci a vývoj UI.

## **Klíčová slova**

Člověk, etika, etické principy, myšlení, odpovědnost, umělá inteligence, vědomí.

## **Abstract**

The thesis focuses on the ethics of artificial intelligence (AI). It is organized into three main parts. The first part describes the basic concepts. There are explained in particular the basic principles of algorithms and the functioning of AI from a technical point of view. In the next part, the thesis uses texts by J.R. Searle and J. Peregrine to compare AI and human thinking from a philosophical perspective. From this section it is concluded that AI is a machine and as such has no moral responsibility. The moral responsibility in working with AI lies with humans, that is, the developers of AI and its users. This is addressed in the third part of the thesis, where various current moral issues are presented and basic ethical principles for AI work and development are established based on utilitarian and deontological orientations.

## **Keywords**

Human, ethics, ethical principles, thinking, responsibility, artificial intelligence, consciousness.

**Počet znaků** (včetně mezer): 142 425

## **Poděkování**

Na tomto místě bych velice ráda poděkovala vedoucímu práce Mgr. Ing. Prokopu Sousedíkovi, Ph.D. za konzultace při výběru tématu a obsahu práce.

Současně bych také chtěla poděkovat přítelkyni Veronice a sestře Alžbětě za pomoc s korekturou textu.

A také moc děkuji své rodině a blízkým za jejich podporu.

# Obsah

Úvod .....	7
1    Základní pojmy.....	9
1.1    Umělá inteligence.....	9
1.2    Etika .....	11
1.3    Základní představení systémů UI.....	11
2    Vývoj umělé inteligence a její vlastnosti .....	16
2.1    Historický vývoj analytické filozofie.....	16
2.2    Charakteristiky umělé inteligence .....	22
2.3    Závěr.....	31
3    Práce se systémy umělé inteligence .....	34
3.1    Mravní zodpovědnost výrobce a zadavatele UI .....	36
3.2    Mravní zodpovědnost uživatele UI .....	50
Závěr .....	64
Seznam literatury .....	67

# Úvod

Tématem diplomové práce je problematika etiky a umělé inteligence. Práce se zabývá morálními aspekty při práci s umělou inteligencí (dále také „UI“).

Téma jsem si vybrala pro jeho aktuálnost. Mělo by se dostat do popředí pozornosti, jelikož se dostáváme do bodu, kdy rychlý vývoj systémů umělé inteligence předbíhá morální zásady práce s nimi. Dalším bodem, který mi napomohl při výběru tématu, je propojení technické a humanitní sféry, tedy konkrétně sféry informačních technologií a oboru etiky, o který se zajímám. V technických oborech jsou dopady na ostatní směry, ale především humanitní sféru, často opomíjené. To je pozůstatek Descartova dualismu, který oddělil fyzikální svět a svět duševní, jinak svět mysli. Po oddělení těchto dvou částí se duševní část dostala zcela do ústraní. Díky rozvoji vědy se pak vše začalo opírat o stránku fyzikální a stránku důkazů. V současné době už se setkáváme se znovunalézáním obou směrů a potřeba jejich propojování vzrůstá.

Diplomová práce je rozdělena na tři části. V první jsou představeny základní pojmy uvozující následující části práce. Zaměřuje se na základní pojmy etiky a typy umělé inteligence. V této části jsou také vybrány a popsány základní systémy umělé inteligence. Výsledkem první kapitoly je stanovení otázek, kterými se budeme zabývat v následující kapitole.

Jedná se o tyto tři základní otázky:

*„Může umělá inteligence myslet?“*

*„Může umělá inteligence nést zodpovědnost za své činy?“*

*„Má umělá inteligence vědomí?“*

Cílem druhé části práce je nalézt odpovědi na tyto tři zásadní otázky, a to za pomoci využití zdrojů vybraných filozofů, kteří se UI ve svých pracích věnují. Jedná se o John R. Searleho, Alana M. Turinga, Friedricha L. G. Fregeho, Rudolfa Carnapa, Daniela Dennetta a další. Na konci druhé kapitoly dojdeme k závěru, že umělá inteligence není autonomní, nemyslí jako člověk, nemá vědomí a nenese odpovědnost za své činy, tudíž nemůže eticky jednat. Tento závěr nás uvádí do třetí části, kde se zaměříme na aktéry etického jednání, kterými jsou na jedné straně zadavatelé a výrobci UI a na straně druhé její uživatelé.

Třetí část se tedy bude zabývat prací se systémy umělé inteligence. Budeme se zde opírat o doposud navržené etické principy a pokusíme se na základě zpracovaných informací tyto principy představit na konkrétních příkladech a odvětvích UI. Cílem této

práce je též ověření stávajících principů a pokud možno jejich doplnění o nové. Stávající obecně platné morální principy jsou tyto: princip beneficence, princip non-maleficence, princip autonomie, princip férovosti a princip transparentnosti.

Třetí část práce bude rozdělena na dvě podkapitoly: „Mravní zodpovědnost výrobce a zadavatele UI“ a „Mravní zodpovědnost uživatele UI“. V první podkapitole se budeme zabývat morálními zásadami tématu ochrany a soukromí dat, rovnosti a nediskriminace, bezpečnosti a ochrany před zneužitím systémů UI, zodpovědnosti a důsledky na straně vývoje těchto systémů, dopadem na udržitelnost a také na společenskou prospěšnost.

Ve druhé podkapitole se pak zaměříme na koncového uživatele, který využívá systémy UI. Většina lidí už s nimi úzce pracuje jak při výkonu povolání, při práci ve škole, tak i ve svém soukromém životě. Používáme chytré telefony, chytré hodinky, roboty a chytré aplikace. Dalšími koncovými uživateli jsou lidé z odborné veřejnosti, kteří umělou inteligenci využívají při výkonu svého povolání, jako jsou například medicí, armáda, policisté, politikové, ekonomové, ale i právníci a další. Tato podkapitola je zaměřena na práci s daty ze strany uživatele, ovlivňování, manipulaci a zneužívání UI, závislost a zodpovědnost a v konečné fázi i etickou vzdělanost a uvědomělost.

Téma etiky a umělé inteligence je velice rozsáhlé a každá kapitola by mohla vydat na samostatnou práci. Úkol, který jsem si v práci stanovila, je toto téma především představit, pokusit se propojit sféru etiky a této novodobé technologie, dále poukázat na zásadní problematiku tohoto tématu a nastínit její možná řešení stanovením etických zásad vycházejících z etických principů.



# 1 Základní pojmy

V úvodní kapitole diplomové práce si představíme základní pojmy dotčeného tématu a jejich odkaz na následnou práci. Tématem diplomové práce je etika a umělá inteligence. V první řadě se tedy zaměříme na definování inteligence a umělé inteligence jako takové a poté si představíme teoretický rámec etiky, který je s danou tematikou spojen.

## 1.1 Umělá inteligence

Umělá inteligence (UI) nebo také v anglickém překladu *artificial intelligence (AI)* je spojení dvou slov, z nichž samostatné slovo inteligence se vztahuje výhradně k člověku. Za zmínku stojí i významový překlad slova *artificial*, který ve vztahu k chování znamená – strojený, vyumělkovaný, falešný.<sup>1</sup> Už samotný překlad napovídá, že se jedná o napodobení, nikoliv autonomní systém. Dle Imricha Ruisela se vědci při vývoji umělé inteligence zabývají imitací, tedy napodobením biologicky předložené inteligence, a to především lidské.<sup>2</sup> Tedy základní pravidla pro fungování systému umělé inteligence nevytváří UI sama, ale vytváří je člověk.

V této práci se nebudeme hlouběji zabývat inteligencí jako takovou, ale je potřeba si zmínit alespoň základní poznatky ohledně dané tematiky, které s naším tématem – morálním aspektem – souvisejí. Lidskou inteligencí se zabývá především psychologie, a to zejména vývojová psychologie člověka. Lidskou inteligenci můžeme rozdělit do několika druhů. Jedním příkladem je dělení podle amerického vývojového psychologa Howarda Gardnera, který rozlišuje inteligenci jazykovou, hudební, matematicko-logickou, prostorovou, tělesnou a kinestetickou, intrapersonální, interpersonální a přírodní.<sup>3</sup> Další převládající typy inteligencí jsou abstraktní, praktická, emoční, sociální a morální.<sup>4</sup> Z těchto základních uvedených typů je zřejmé, že některé druhy lze imitovat či naučit umělou inteligenci snáze. A vzhledem k tomu, že vývojáři umělé inteligence vycházejí z inteligence lidské napodobují ji, je také patrné, že se umělá inteligence stroje bude té lidské podobat. Například matematická inteligence nebude takový problém jako inteligence emoční či morální.

---

<sup>1</sup> *Anglicko-český, česko-anglický velký slovník: - nejen pro překladatele*. 2. vyd. [Brno]: Lingea, c2007.

<sup>2</sup> Srov. RUISEL, Imrich. *Základy psychologie inteligence*. Praha: Portál, 2000. 134.

<sup>3</sup> Srov. GARDNER, Howard. *Frames of mind: the theory of multiple intelligences*. New York: BasicBooks, 1983. xxxii. 17–26.

<sup>4</sup> Srov. RUISEL, Imrich. *Základy psychologie inteligence*. Praha: Portál, 2000. 41.

Uvedme si také jednu z definic umělé inteligence. Pro účely práce byla vybrána tato: „*Intelligence je schopnost dělat správnou věc ve správný čas. Je to schopnost reagovat na příležitosti a výzvy, které přináší kontext.*“<sup>5</sup> V této definici je uvedeno, že se jedná o schopnost dělat správnou věc, to nás přivádí k etické otázce jednání člověka a jeho podobností s prací umělé inteligence. Touto otázkou se budeme zabývat v následujících kapitolách. Další definicí, která odkazuje na podobnost lidského chování a chování systémů umělé inteligence, je definice Marvinů Minského z roku 1967 (uvedena v knize Ivana Zelinky: *Umělá inteligence: hrozba nebo naděje?*): „*Umělá inteligence je věda o vytváření strojů nebo systémů, které budou při řešení určitého úkolu užívat takového postupu, který by lidé považovali za projev jeho inteligence. To znamená, jako kdyby ho dělal člověk.*“<sup>6</sup>

Mezi další osobnosti, které se věnovaly definování umělé inteligence, patří například Alan Turing a Isaac Asimov. Alan Turing postavil svou definici na experimentu, kde osoba komunikuje přes počítač se dvěma dalšími aktéry, z nichž jeden je člověk a druhý počítač. Dle jeho definice se práce počítače dá považovat za umělou inteligenci, pokud člověk nerozezná, co odpovídá počítač a co člověk. Tento experiment má však několik nedostatků, a to především ve volbě osob.<sup>7</sup> Výše zmíněný Isaac Asimov pak ve své literatuře uvádí tři zákony robotiky.<sup>8</sup> Tyto zákony však narážejí na to (jak bude popsáno v další kapitole), že stroje nepřemýšlí a nemají odpovědnost, a tedy zákony by měly být stanoveny pro lidi, nikoliv pro stroje.

---

<sup>5</sup> BRYSON J. Joanna. „The artificial intelligence of the ethics of artificial intelligence. An Introductory Overview for Law and Regulation“ In *The Oxford Handbook of Ethics of AI*, edited by DUBBER, D. Markus, PASQUALE, Frank, DAS, Sunit. Oxford University Press, 2020. 4.

<sup>6</sup> In: ZELINKA, Ivan. *Umělá inteligence: hrozba nebo naděje?* Praha: BEN-technická literatura, 2003. 20–21.

<sup>7</sup> Tamtéž, 26.

<sup>8</sup> Tři zákony robotiky sepsal Isaac Asimov ve své knize: „*Já, robot*“. Zákony robotiky zní takto:

1. Robot nesmí ublížit člověku nebo svou nečinností dopustit, aby člověku bylo ublíženo.

2. Robot musí uposlechnout příkazů člověka, kromě případů, kdy tyto příkazy jsou v rozporu s prvním zákonem.

3. Robot musí sám sebe chránit před zničením, kromě případů, kdy tato ochrana je v rozporu s prvním nebo druhým zákonem.

In: ASIMOV, Isaac. *Já, robot*. Přeložil Oldřich ČERNÝ, přeložil Alexandr KRAMER, přeložil Zuzana MEYEROVÁ. Trifid (Triton). Praha: Triton, c2012.

## 1.2 Etika

V kapitole 1.1. jsme se zabývali pojmem umělá inteligence. V této části si představíme základní pojmy spojené s etikou, které budou sloužit jako úvod do následující práce.

Pro představu novodobého vědeckého pojetí etiky uveďme definici A. Arzenbachera: „*Etika je vědecká (filozofická nebo teologická) disciplína, jejímž předmětem jsou správná pravidla (normy) lidského jednání neboli lidské praxe.*“<sup>9</sup> Z definice je patrné, že se etika zabývá morálním aspektem lidského jednání a lidské praxe. Z tohoto pohledu je etika důležitým nástrojem, který nám pomáhá určovat, co je správné a co je špatné, jak bychom měli jednat a čeho bychom se měli vyvarovat. Etika umožňuje reflektovat naše morální hodnoty a pomáhá nám žít v souladu s nimi. Etikou jednání se zabýval například Immanuel Kant ve své knize *Základy metafyziky mravů*<sup>10</sup> a také Aristoteles v knize *Etika Nikomachova*.<sup>11</sup> Kant popisuje jako nejvyšší princip mravnosti autonomii vůle. Autonomii vůle pak popisuje takto: „*Autonomie vůle je povaha vůle, díky níž je tato vůle sama sobě zákonem (nezávisle na veškeré povaze předmětů chtění).*“<sup>12</sup> Autonomii pak popisuje také jako podmínku pro svobodnou vůli člověka.<sup>13</sup> Aristoteles se ve své knize zabývá otázkou odpovědnosti a jako základní podmínky pro odpovědné chování uvádí dobrovolnost a vědomost jednání.<sup>14</sup> Závěrem tedy lze konstatovat, že aby bylo jednání etické, musí být autonomní, dobrovolné a vědomé. Tématem této práce je nalézt odpovědi na otázky, kdo je mravně odpovědný za jednání umělé inteligence, zda člověk, nebo stroj.

## 1.3 Základní představení systémů UI

V této práci je kladen důraz na jeden z hlavních třech pilířů etiky, a to etiku aplikovanou.<sup>15</sup> Aplikovaná etika se zabývá jednáním a praxí člověka v různých oborech, jako jsou medicína, právo, politika a mnoho dalších oborů. Abychom se mohli aplikovanou etikou zabývat, nestačí pouze znalost norem a jazyka etiky, tedy analytická a základní část, ale také podrobnější znalost daného oboru. Například nelze morálně

---

<sup>9</sup> ANZENBACHER, Arno. *Křesťanská sociální etika: úvod a principy*. 2. vydání. Přeložil Karel ŠPRUNK. Brno: CDK, Centrum pro studium demokracie a kultury, 2015.9.

<sup>10</sup> KANT, Immanuel. *Základy metafyziky mravů*. 2. vyd. Filozofické dědictví. Praha: Svoboda, 1990.

<sup>11</sup> ARISTOTELÉS. *Etika Nikomachova*. Páté nezměněné vydání. Praha: Rezek, 2021.

<sup>12</sup> KANT, Immanuel. *Základy metafyziky mravů*. 2. vyd. Filozofické dědictví. Praha: Svoboda, 1990. 59.

<sup>13</sup> Tamtéž, 79.

<sup>14</sup> Srov. ARISTOTELES. *Etika Nikomachova*. Páté nezměněné vydání. Praha: Rezek, 2021. 1110.

<sup>15</sup> Aplikovaná etika je jedním ze čtyř pilířů etické disciplíny. Jako jediná zasahuje do praxe. Ostatní pilíře etiky (deskriptivní, normativní a metaetika) jsou teoretické.

In: THOMPSON, Mel. *Přehled etiky*. Vyd. 1. Praha: Portál, 2004. 167.

hodnotit jednání politických aktérů pouze na základě znalostí analytické a normativní etiky, ale je nutnou součástí hodnocení znát i kontexty a konkrétní problémy politické sféry. Především ve vývoji a fungování umělé inteligence je důležité, aby byl kladen důraz na propojení a vzájemné doplňování oborů humanitních a technických.

Tato práce se bude zabývat především morálním jednáním a s tím spojenou problematikou umělé inteligence. Pro toto hodnocení je však důležité se seznámit i s technickým fungováním umělé inteligence. Proto si v této kapitole uvedeme základní algoritmy a technická řešení UI.

UI funguje na principu strojového učení. Jedná se o poskytování velkého množství dat a vytvoření takového algoritmu, který stroj přiměje data zpracovávat a vyhledávat si další, jež jsou pro jeho funkci přínosná. UI rozdělujeme na softwarovou a hardwarovou. Mezi hardwarovou UI zařazujeme především robotiku, která se v současnosti stala rozsáhlým samostatným oborem. Mezi softwarovou UI řadíme programy a funkce s využitím UI. UI můžeme také dělit na silnou a slabou. Se slabou umělou inteligencí se setkáváme v běžném životě, je naprogramována na určitý typ úkolu. Naopak v silné umělé inteligenci jde o naprogramování takového stroje, který by myslel jako člověk a má vědomí. V případě silné umělé inteligence se jedná o hypotetický konstrukt. Zatím nejsou vyjasněny otázky, zda je vůbec možné, aby umělá inteligence myslela či měla vědomí a odpovědnost. Těmito otázkami se zabývá diplomová práce v kapitole 2.1.1.

Jak bylo popsáno výše, pro etické hodnocení daného tématu je potřeba znalost konkrétního oboru. Tato diplomová práce se zabývá na jedné straně filozofickými otázkami, ale na druhé straně je spojena s umělou inteligencí jako technickým oborem. Proto budou níže uvedeny a popsány základní používané algoritmy a principy umělé inteligence, jako je například fuzzy logika, neuronové sítě, evoluční sítě, umělý život a systém GAN.

### *Fuzzy logika*

Fuzzy logika neboli „mlhavá“ logika vděčí za svůj vznik L. A. Zadehovi. Nejpřesněji je možné toto řešení popsat na Aristotelově výroku: „*Systém je víc nežli pouhý souhrn jeho částí.*“<sup>16</sup> Do doby představení fuzzy logiky byly matematické algoritmy vyvíjeny na systému 1 a 0 (pravda, nepravda). Tímto binárním kódem se daly popsat takové obory, u kterých bylo zřejmé, co k jakému číslu přiřadit, zejména se jedná

---

<sup>16</sup> In. ZELINKA, Ivan. *Umělá inteligence: hrozba nebo naděje?* Praha: BEN – technická literatura, 2003. 55.

o přírodní vědy. To bylo však nevyužitelné pro obory, jako jsou například ekonomie, ekologie, sociologie nebo psychologie. Zmíněné obory jsou komplexnější a nelze z nich vytvořit závěry pouze z pravd a nepravd. Pro tento případ přišel L. A. Zadeh s matematickými algoritmy, které pracují s celým intervalem  $\langle 0,1 \rangle$ . Můžeme si to představit jako hodnoticí formulář, kde už nyní nemáme pouze striktní odpovědi ano/ne, pořád/vůbec, ale jsou zde i mezičlánky – ano – spíše ano – nevím – spíše ne – ne, nebo často – méně často – skoro ne – vůbec ne. Výsledná hodnota formuláře bude pak více odpovídat realitě, i když bude „zamlžená“ a nestanoví nám jen rigidní odpověď.<sup>17</sup>

### *Neuronové sítě*

Jako další výrazný matematický model UI je používána neuronová síť. Jak už název napovídá, jedná se o síť, která byla vyvíjena na základě podobnosti s biologickou neuronovou sítí. Historie tohoto vývoje spadá až do 40. let 20. století, kdy W. H. Pitts vytvořil první model neuronu, který se používá dodnes. Dalším průlomovým bodem bylo vytvoření vícevrstevných sítí, které řeší problémy i nelineárně. Za tímto úspěchem stojí vývojáři D. Rumelhart, G. Hinton, a R. Williams, kteří své poznatky zaznamenali v práci *Learning Internal Representation by Error Propagation*.<sup>18</sup> Princip dnes používaných vícevrstevných sítí lze představit následovně. První vrstva neuronů je vstupní vrstva (větvící), další je vrstva skrytá a poslední vrstva výstupní. Vektory mezi sítěmi mají danou váhu, což je veličina, která stanovuje význam vstupu pro daný neuron, nikoliv pro řešený problém. Samotná funkce neuronové sítě se pak dělí na fázi adaptační (vybavovací) a aktivační (učící). Při adaptační fázi je použita trénovací množina, která stanovuje vstupní i výstupní vektory. Vstupní vektor se přepočítá přes nastavený algoritmus a přiřadí mu určité váhy. Výstupní vektor se pak porovná s originálem a stanoví se odchylka. Při větší odchylce se celý proces opakuje a stanovují se jiné váhy, dokud není v nejlepším případě naučený výsledek totožný s originálem. Zde také rozlišujeme učení s učitelem a bez učitele. V případě učení s učitelem máme zadány vstupní i výstupní vektory, v opačném případě pracujeme pouze se vstupními vektory.<sup>19</sup> Celý problém si lze představit na příkladu. Stanovme několik tříd se svými zástupci – kolo, auta, lodě apod. Pro trénovací množinu vybereme část informací.

---

<sup>17</sup> Srov. ZELINKA, Ivan. *Umělá inteligence: hrozba nebo naděje?* Praha: BEN – technická literatura, 2003. 55.

<sup>18</sup> In: ZELINKA, Ivan. *Umělá inteligence: hrozba nebo naděje?* Praha: BEN – technická literatura, 2003. 39.

<sup>19</sup> Srov. ZELINKA, Ivan. *Umělá inteligence: hrozba nebo naděje?* Praha: BEN – technická literatura, 2003. 40.

V každé této třídě bude vzorový vektor, podle kterého se daná síť bude učit a stanovovat váhy. Tedy námi požadovaný vstup a výstup: pick-up -> „jedná se o auto“. V případě, že síť má výstup odlišný, vrátí se na začátek a upraví své hodnoty, dokud nedojde ke stejnému výsledku. Po učicí fázi se síť aplikuje na všechny informace.

Neuronová síť se tedy od běžného programu liší tím, že se nejedná o jeden algoritmus, nýbrž o více algoritmů, kde rozhodovací schopnost a výstup není dán klasickým programovacím jazykem (if, then, go, to...), ale souhrnem neuronů a vah mezi nimi. Díky tomuto systému se síť může učit a adaptovat i na další problémy, zatímco v běžném programu je nutné každou změnu či adaptaci upravit a naprogramovat. Je to tedy komplexnější a provázanější řešení než klasické počítačové programy.

Neuronové sítě lze využít ve vojenských aplikacích pro identifikaci radarových či sonarových signálů, pro predikci chování – vývoj burzy apod., dále pro klasifikaci, optimalizaci nebo filtraci.<sup>20</sup>

#### *Evoluční a genetický algoritmus*

Evoluční algoritmy byly naprogramovány na základě Darwinovy evoluční teorie. Jedná se o vývoj stále nových a nových výpočtů, z nichž jsou vybírány ty nejlepší, a s těmi se dále pracuje. Původně byly tyto algoritmy používány především v genetice a kopírovaly genetické procesy lidí. Nyní se tento typ algoritmu používá i v dalších sférách, především ve výpočetní technice, například k optimalizaci trajektorie robota či optimalizaci chodu leteckých motorů Boeing. Díky tomuto systému lze simulovat i celý evoluční systém a vytvářet tak uměle životní prostředí, tzv. „umělý život“<sup>21</sup>. Na bázi evolučního algoritmu lze představit další typ algoritmu, který kopíruje biologické procesy člověka, a to genetický algoritmus. Jedná se o algoritmus, který vytváří simulace lidského DNA a následný vývoj nových zdravých jedinců. Přestože se tento proces v současnosti ještě nevyužívá, je potřeba si uvědomit, že tento „výběr“ představuje velký etický problém.

#### *GAN*

GAN neboli generativní kontradiktorní sítě (Generative Adversarial Network) představuje novodobou metodu tvorby umělé inteligence založenou na dvou algoritmech, které jsou postaveny proti sobě a tímto principem se zdokonalují a rozvíjí

---

<sup>20</sup> Srov. ZELINKA, Ivan. *Umělá inteligence: hrozba nebo naděje?* Praha: BEN – technická literatura, 2003. 40.

<sup>21</sup> Tamtéž, 65.

ve svém učení.<sup>22</sup> Tato metoda se používá při práci s obrazovým materiálem. Zatímco se jeden algoritmus snaží daný obraz či fotografii napodobit, druhý algoritmus vyznačuje rozlišnosti oproti originálu a vrací obraz zpět. Tato metoda je velmi efektivní. Pracuje s algoritmem vývoje a kontroly a tím dochází k rychlejší práci a zároveň k optimálnímu řešení výsledku. Tato metodika však už byla i zneužita pro kybernetické útoky, kde došlo k vytvoření nové identity člověka, jemuž byly vytvořeny fotografie a otisky prstů, což bylo zneužito v bankovním systému. V současnosti nachází GAN uplatnění především při identifikaci osob, ale také v uměleckém světě pro odhalování plagiátů. Má potenciál otevřít nové možnosti v oblasti architektury a tvorby modelů založených na 2D snímcích.<sup>23</sup>

Po představení základních používaných algoritmů umělé inteligence si lze všimnout, že velká část z nich je simulací přírodních procesů, ať už se jedná o biologické procesy v těle člověka, tak o celkové evoluční systémy. I přes tento vývoj však nelze zapomínat, že jde pouze o strojové schéma, nikoliv o „postavení“ nového člověka na základě těchto principů.

---

<sup>22</sup> Srov. BROWNLEE, Jason. „A Gentle Introduction to Generative Adversarial Networks (GANs). Machine Learning Mastery.“ In: *Generative Adversarial Network* [online]. 2019. [cit. 2024-02-11]. Dostupné z: <https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/>

<sup>23</sup> Srov. RUSANEN, Anna-Mari. „Chapter 1: What is AI ethics?“ In: *Ethics of ai* [online]. [cit. 2022-06-10]. Dostupné z: <https://ethics-of-ai.mooc.fi/chapter-1/1-a-guide-to-ai-ethics>

## 2 Vývoj umělé inteligence a její vlastnosti

V této kapitole se v první části seznámíme s historickými milníky, které vedou k otázkám myšlení a vývoje této otázky v analytické filozofii. To nám bude následně sloužit pro odpovědi na otázky myšlení a fungování strojů. V druhé části této kapitoly si představíme základní charakteristiky umělé inteligence. Zaměříme se na otázku mysli, autonomie a vědomí. Cílem této kapitoly jsou odpovědi na tyto základní filozofické otázky k tématu etiky a umělé inteligence:

„*Mohou stroje myslet?*“

„*Mohou být stroje zodpovědné?*“

„*Mohou mít stroje vědomí?*“

Zodpovězení těchto otázek se stává čím dál naléhavějším, protože vývoj UI je nezadržitelný a nové technologie se stále více a rychleji integrují do našich životů. Před námi je tedy důležitá úloha zkoumání a řešení těchto otázek, aby bylo možné řídit vývoj a aplikaci umělé inteligence tak, aby byla v souladu s etickými hodnotami a zájmy.

### 2.1 Historický vývoj analytické filozofie

Základní kořeny otázky myšlení a umělé inteligence spadají až do počátků 20. století. V té době se nejednalo o konkrétní řešení uvedené problematiky, ale o velmi důležité propojení filozofie a matematické logiky, čímž se začali zabývat analytičtí filozofové. Ve svých počátcích se ale prvotně nezaměřovali ani na problematiku lidského či strojového myšlení. Místo toho se orientovali více na metody zkoumání a poznávání a také na vymezení klasické filozofie a jejího smyslu.<sup>24</sup> Později se však na světlo dostaly stroje a první počítače a začaly se otevírat otázky ohledně jejich práce a jejich možností.

Prvotně se tedy analytičtí filozofové zaměřovali na metodu zkoumání, která je postavena především na zkušenostech – empirii, významu přirozeného jazyka a formální logice.<sup>25</sup> Jaroslav Peregrin definoval analytickou filozofii takto: „pokus o promýšlení tradičních filozofických problémů matematickým rozumem dvacátého století.“<sup>26</sup> Jaroslav Peregrin ve své knize *Kapitoly z analytické filozofie* uvedl myšlenky klasických analytických filozofů, jako byli například Gottlob Frege, Bertrand Russell,

---

<sup>24</sup> Srov. PEREGRIN, Jaroslav. *Kapitoly z analytické filozofie*. Praha: Filosofía, 2005. 14–16.

<sup>25</sup> Tamtéž, 16-17.

<sup>26</sup> PEREGRIN, Jaroslav. *Kapitoly z analytické filozofie*. Praha: Filosofía, 2005. 17.



Rudolf Carnap či Ludwig Wittgenstein,<sup>27</sup> které lze využít pro popis poznávání filozofických aspektů fungování umělé inteligence pomocí analytické logiky.

Prvním zmíněným je významný filozof, logik a matematik Gottlob Frege, který se svým díle hlouběji zabýval otázkami matematické logiky a soustředil se na pravdivost matematických důkazů. Ve své knize *Pojmové písmo* se dostal i k problematice dokazování versus myšlení, tedy k rozdílu mezi logikou a psychologií. Zatímco v matematice se zabýváme dokazováním a odůvodňováním, mechanika duševních procesů probíhá jinak. V logice jde především o to, jak daný výrok dokázat. K tomu chtěl Frege dojít stanovením elementárních vzorců, z nichž budou poskládány důkazy, poté se bude rozhodovat o pravdě či nepravdě daného výroku. Toto vyplývání je záležitostí faktu. Psychologie na druhou stranu zkoumá proces myšlení, tedy jak se člověk k danému výroku fakticky dopracuje.<sup>28</sup> Fregeovo dílo se primárně nezabývalo rozdílem mezi logikou a psychologií. Jeho práce však předkládá další otázky pro následné zkoumání a jeho úvahy lze aplikovat na fungování umělé inteligence. UI funguje na základě provádění příkazů, které jsou zadávány podle definovaných pravidel a vzorců. Pokud bychom však definici „myšlení“ přijali z Fregeova výkladu, tedy že se jedná o proces zkoumaný psychologií, docházím dle Fregeovy teorie k závěru, že UI funguje na bázi logiky a matematiky a nelze říct, že „myslí“ jako člověk.

Další skupinou, která představuje myšlení analytických filozofů, je skupina okolo Vídeňského kruhu.<sup>29</sup> Tato skupina byla dosti radikální ve svých názorech ohledně filozofických problémů a přírodních věd. V knize *Vědecký světový názor* bylo uvedeno, že každý problém má řešení. Tento radikální směr navrhoval rozdělovat tvrzení a problémy na základě logické analýzy a empirie. Pokud tvrzení nebylo možné rozložit logickou analýzou, bylo považováno za nesmyslné. Pokud bylo možné ho dále rozložit, přesunulo se k určené přírodní vědě.<sup>30</sup> Carnap ve své knize *Logická výstavba světa* popisuje poznávání pomocí logické analýzy, jež zakládá na smyslových datech aneb poznání na základě smyslových vjemů. Pokud je poznání prvotní, stává se z něj elementární prožitek. Elementární prožitky nemusí být totožné, ale souhrn prožitků

---

<sup>27</sup> Srov. PEREGRIN, Jaroslav. *Kapitoly z analytické filosofie*. Praha: Filosofia, 2005. 37.

<sup>28</sup> Tamtéž, 37.

<sup>29</sup> Vídeňský kruh vznikl v roce 1929 a jedná se o skupinu vědců, která chce tradiční filozofii překonat moderní vědou.

In: PEREGRIN, Jaroslav. *Kapitoly z analytické filosofie*. Praha: Filosofia, 2005. 95.

<sup>30</sup> Tamtéž, 96.

může vyvolat pouze podobnost, to Carnap popisuje jako „vzpomínku podobnosti“<sup>31</sup>. Na těchto dvou příkladech vidíme, jak postupovali v poznávání analytičtí filozofové, a také to lze porovnat s prací UI.

V druhé polovině 20. století se na scéně objevily počítače a jejich fungování a plnění úkonů začalo vzbuzovat otázky ohledně podobnosti práce počítačů a lidského jednání a myšlení. Tyto otázky se staly nedílnou součástí další odnože analytické filozofie. Mezi hlavní osobnosti tohoto tématu patří Alan Turing a John R. Searle.<sup>32</sup>

Alan Turing je veřejně známý zejména pro svůj přínos v kryptografii a kybernetice, ale také pro teorii mysli a takzvaný „Turingův test“. Turing se narodil v roce 1912 v Londýně. V roce 1935 dokončil studia matematiky na King's College v Cambridgi a v roce 1937 publikoval svůj článek *On Computable Numbers, with an Application to the Entscheidungsproblem (O vypočitatelných číslech)*, kde uvedl základy informatiky. Od roku 1938 působil v britské armádě, kde sloužil pro kryptografické oddělení a přispěl k rozluštění kódu šifrovacího stroje Enigma.<sup>33</sup> Pro Turingovu teorii mysli je stěžejní rok 1950, kdy publikoval článek *Computing Machinery and Intelligence*. V tomto článku byla poprvé rozebrána důležitá otázka pro nadcházející dobu informatiky, a sice „zda mohou stroje myslet?“. K zodpovězení této otázky vymyslel Turing test, později nazývaný právě jako Turingův test. Spočíval ve srovnání komunikace lidské bytosti a stroje. Na jedné straně byl zadavatel a na druhé straně byly dvě entity, které se zadavatelem komunikovaly, přičemž jedna entita byla lidská bytost a druhá stroj – počítač. Dle odpovědí zadavatel určil, zda odpovídá stroj nebo člověk. Pokud by teoreticky stroj oklamal zadavatele po celou dobu a nebylo by možné rozpoznat počítač od člověka, bylo by pak možné říct, že počítače myslí? Turing přeformuloval otázku myšlení na otázku: „Může počítač hrát úspěšně roli člověka?“ Turing myšlení prisuzoval pouze živým organismům, avšak v roli počítačů uvažoval o rozšíření slova myšlení i na neživé organismy. Turingův pokus a nastolení otázky o myšlení strojů rozpoutalo nové bádání v oblasti analytické filozofie a kognitivní vědy. V roce 1954 Turing spáchal sebevraždu a dalšímu rozvoji v jeho pojetí této otázky nedošlo. I tak položil základní kámen pro další bádání.<sup>34</sup>

---

<sup>31</sup> Srov. PEREGRIN, Jaroslav. *Kapitoly z analytické filozofie*. Praha: Filosofía, 2005. 37.

<sup>32</sup> Tamtéž, 231.

<sup>33</sup> Srov. TVRDÝ, Filip. *Turingův test, Filosofické aspekty umělé inteligence*. Olomouc, 2011. Disertační práce (PhD.). Univerzita Palackého v Olomouci, Filosofická fakulta, Katedra filozofie, 12.

<sup>34</sup> Srov. PEREGRIN, Jaroslav. *Kapitoly z analytické filozofie*. Praha: Filosofía, 2005. 231.

Turingův test nejen položil kámen pro další zkoumání otázky myšlení, ale také vyvolal spoustu reakcí. Jaroslav Peregrin ve své knize *Kapitoly z analytické filosofie* uvádí jako oponenty Turingova kritéria myšlení například Johna Searla, Neda Blocka či Rogera Penrose. Dále se zmiňuje o filozofovi Danielu Denettovi, který částečně Turingův test akceptuje.<sup>35</sup> Všichni zmínění se v reakci na Turinga zabývali otázkou myšlení strojů.

Jako jeden z hlavních kritiků Turingova testu a také průkopník filozofie v otázce myšlení a umělé inteligence je John Searle, americký profesor filozofie na Kalifornské univerzitě v Berkley. Jeho pojetí vztahu mysli a mozku rozhodně stojí za zmínku. Filozof ve svých úvahách popisuje striktní Descartův dualismus. Právě ten byl jedním z hlavních průlomů ve filozofii, kdy spojení duše a těla bylo rozděleno. Dle Descartova dualismu stojí myšlenky a city na jedné straně a jsou popisovány jako něco subjektivního, vědomého a nemateriálního. Na opačné straně stojí fyzikální fakta, která jsou hmotná, materiální a dokazatelná a dle Searla mají kauzální interakci s okolím. Descarte svým dualismem oddělil fyzickou a duševní stránku. S vývojem společnosti po fyzikální stránce a s rozvojem přírodních vět se stala fyzická strana primární a ta duševní se dostala do pozadí, popřípadě byla přímo popírána. Searle však oponuje a svou prací dokazuje, že není možné duševní stranu neboli stranu mysli popřít a zaměřit se pouze na stranu materialistickou.<sup>36</sup> Své tvrzení opírá o čtyři důležité mentální vlastnosti:

1) Vědomí

*„Vědomí je totiž ústředním faktem lidské existence, protože bez vědomí by nebyly možné všechny ostatní, typicky lidské aspekty naší existence: jazyk, láska, humor atd.“<sup>37</sup>*

2) Intencionalita

*„Intencionalitou jsou naše mentální stavy zaměřeny na předměty a stavy věcí světa, odlišné od stavů mentálních, které se předmětů a stavů věcí týkají či na ně poukazují. Ostatně „intencionalita“ se netýká jen intencí či záměrů, ale také přesvědčení, přání, nadějí, obav, lásky, nenávisti, žádosti, hnusu, studu, hrdosti, vzteku, pobavenosti a všech ostatních mentálních stavů (ať už vědomých nebo nevědomých), které se*

---

<sup>35</sup> Srov. PEREGRIN, Jaroslav. *Kapitoly z analytické filosofie*. Praha: Filosofia, 2005. 237.

<sup>36</sup> Srov. SEARLE, John R. *Mysl, mozek a věda*. Přeložil Marek NEKULA. Praha: Mladá fronta, 1994. Váhy (Mladá fronta). 15–16.

<sup>37</sup> Tamtéž, 15–16.

vztahují na svět mimo sebe a které se jím také zabývají.“<sup>38</sup> Jinými slovy lze intencionalitu popsat dle fenomenologického výkladu, že se jedná o „vztažnost“ ke světu aneb že každý člověk se k něčemu vztahuje. Při svém jednání se vztahuje k místu, času, okolí, lidem. Vztažnost dle fenomenologie končí jediné smrtí.<sup>39</sup>

### 3) Subjektivita mentálních jevů

Jedná se o subjektivní pocity a vyjadřování každého individuálního člověka. Searle přikládá subjektivitu mentálnímu světu, zatímco vědecký svět ohraničuje objektivními daty.<sup>40</sup> V tomto ohledu se vedlo a vede mnoho diskusí, jelikož disciplíny jako například psychologie se zabývají především subjektivními pocity, a je tedy otázkou, do jaké míry mohou patřit do vědních oborů, pokud nejsou fyzikálně dokazatelné.

### 4) Mentální kauzace

Mentální kauzace je jev, kdy naše subjektivní vnímání a naše pocity a myšlenky ovlivňují okolní svět, tedy že naše chování a jednání kauzálně působí na fyzický svět.<sup>41</sup>

Searl uvádí, že tyto čtyři vlastnosti jsou podstatné pro teorie mysli a vztahu mysli a těla.<sup>42</sup>

Dalším důkazem, kde Searl odporuje dualistickému pojetí mysli a mozku, je tvrzení, že mysl a mozek se zároveň ovlivňují, ale nejsou to dvě odlišné věci.<sup>43</sup> Toto tvrzení Searl opírá o dvě teze.

- 1) „*Mentální fenomény, myslím tím veškeré mentální fenomény – ať už vědomé nebo nevědomé, jsou výsledkem procesů, jež se odehrávají v mozku.*“<sup>44</sup>
- 2) „*Mentální fenomény jsou právě vlastnostmi mozku.*“<sup>45</sup>

Jedním z dalších zmíněných filozofů, který reagoval na Turingův test a jemuž se věnuje Jaroslav Peregrin ve své knize *Kapitoly z analytické filosofie*, byl Ned Block se svým experimentem z roku 1981. Jeho přínosem byla představa počítače, který má

---

<sup>38</sup> Srov. SEARLE, John R. *Mysl, mozek a věda*. Přeložil Marek NEKULA. Praha: Mladá fronta, 1994. Váhy (Mladá fronta). 16.

<sup>39</sup> Srov. HEIDEGGER, Martin a PETŘÍČEK, Miroslav. *Bytí a čas*. Třetí, opravené vydání. Přeložil Ivan CHVATÍK, přeložil Pavel KOUBA, přeložil Jiří NĚMEC. Knihovna novověké tradice a současnosti. Praha: OIKOYMENH, 2018. 287.

<sup>40</sup> Srov. SEARLE, John R. *Mysl, mozek a věda*. Přeložil Marek NEKULA. Praha: Mladá fronta, 1994. Váhy (Mladá fronta). 16–17.

<sup>41</sup> Tamtéž, 16–17.

<sup>42</sup> Tamtéž, 18.

<sup>43</sup> Tamtéž, 21.

<sup>44</sup> Tamtéž, 18.

<sup>45</sup> Tamtéž, 20.

v sobě základní věty ke konverzaci (cca na 1 hodinu konverzace). Během Turingova testu je schopný komunikovat a vést rozhovor svým směrem za využití jeho vět. To ovšem nedokazuje kritérium myšlení, jelikož stroj sám ovlivňuje průběh hovoru dle své kapacity.<sup>46</sup> Pokud tento test aplikujeme do současného světa, představme si novodobé virtuální asistenty – například Siri od firmy Apple. Tito asistenti mají vysokou zásobu vět a vyhledávacích schopností. Vzhledem k jejich naprogramování mohou ovlivňovat uživatele k naprogramovaným účelům, například k reklamní propagaci určitých materiálů.

Jaroslav Peregrin popisuje i Rogera Penrose a jeho argumentaci proti Turingovu testu tím, že myšlení lidí se ne vždy řídí podle daných pravidel, což je u stroje samozřejmé. Stroj se vždy bude řídit podle naprogramovaných a naučených informací. Toto tvrzení vycházelo z Gödlova názoru, že v rámci aritmetiky bude vždy existovat výrok, který je pravdivý, ale nikoliv dokazatelný.<sup>47</sup> Penrose ve shodě se Searlem konstatuje, že „člověk se od stroje liší tím, že dokáže něco více než jenom následovat daná pravidla“<sup>48</sup>.

Další filozof myslí, kterého Peregrin popisuje, je Daniel Dennett. Ten se k problematickému myšlení umělé inteligence vyjadřuje pomocí „Teorie postojů“. Daniel Dennett ve své teorii popisuje, že ke zkoumání okolních entit používáme různé druhy postojů, například fyzikální postoj [physical stance] – tím popisujeme entity, které se řídí podle zákonů fyziky. Dalším postojem je projektový postoj [design stance], zde předvídáme chování entity dle jejího naprogramování. Postoj, kterým nahlížíme na člověka, je pak intencionální [intencional stance], dle jeho přání a přesvědčení pak nahlížíme na jeho chování.<sup>49</sup> Daniel Dennett také reaguje na Rogera Penrose, a to proti jeho postoji „nealgoritmického jednání“. Dennett tvrdí, že i člověk je svým způsobem robot, jelikož funguje na základech fyziky, a ačkoliv je jeho mozek složen z organických součástí, pořád se řídí deterministickými pravidly. Dennett tedy říká: „Nejlepším důvodem pro přesvědčení, že roboti jednoho dne budou moci myslet, je fakt, že my, lidské bytosti myslíme, a přitom i my jsme svého druhu roboti.“<sup>50</sup>

---

<sup>46</sup> Srov. PEREGRIN, Jaroslav. *Kapitoly z analytické filosofie*. Praha: Filosofia, 2005. 237.

<sup>47</sup> Tamtéž, 238.

<sup>48</sup> In: PEREGRIN, Jaroslav. *Kapitoly z analytické filosofie*. Praha: Filosofia, 2005. 238.

<sup>49</sup> Srov. PEREGRIN, Jaroslav. *Kapitoly z analytické filosofie*. Praha: Filosofia, 2005. 239.

<sup>50</sup> In: PEREGRIN, Jaroslav. *Kapitoly z analytické filosofie*. Praha: Filosofia, 2005. 239.

### 2.1.1 Shrnutí

V této kapitole jsme si nastínili vývoj od počátku analytické filozofie až k problematice myšlení strojů, a tedy umělé inteligence. Základním kamenem byl vůbec vznik analytické filozofie a propojení matematické logiky a filozofických problémů. Následně analytičtí filozofové pracovali především na metodách zkoumání, které jsou postaveny, jak bylo uvedeno, na empirii, logice a jazyku. Toto zkoumání následně otvíralo další otázky, a to vztah logiky a psychologie, popřípadě logiky a duševního jednání. Následně se na scéně objevily první stroje a tím se významně rozšířily filozofické úvahy o jejich fungování a práci. Průkopníkem této otázky byl Alan Turing. Představili jsme si také různé reakce na jeho úvahy, které budou sloužit jako základní kámen pro zodpovězení otázky myšlení umělé inteligence, kterou se budeme zabývat v následující kapitole.

## 2.2 Charakteristiky umělé inteligence

### 2.2.1 Práce UI versus lidské myšlení

V předešlé kapitole jsme se seznámili s filozofy, kteří začali ve svém díle propojovat filozofii a matematickou logiku a tím se dostali k otázkám myšlení lidí a poté k myšlení strojů. V této kapitole navážeme na předchozí výklad a budeme ho aplikovat na problematiku umělé inteligence. Jednou z výše uvedených klíčových otázek v oblasti umělé inteligence a jejích etických aspektů, kterou se zabývali zprvu odborníci a postupem času se k ní začala připojovat i široká veřejnost, je otázka myšlení, konkrétně: „Zda může umělá inteligence myslet?“ Otázka myšlení počítačů se spojuje především se silnou umělou inteligencí. Peregrin uvádí ve své knize argumentaci vědců, jako jsou například Herbert Simon a Allen Newell, kteří se domnívají, že umělá inteligence dokáže myslet. Jejich argumentace spočívá v tom, že inteligence není způsobena biologickými či fyzikálními akty, ale je to záležitost pouze fyzikálních operací se symboly.<sup>51</sup>

Hlavním průkopníkem v otázce myšlení a práce strojů byl John Searle. Ten ve své práci konstatoval, že počítač je naprogramován na základě 1 a 0 a tyto základní číslice pak tvoří program, jedná se tedy o řetězec symbolů, zatímco v lidské mysli má tento řetězec symbolů i význam, tedy nejen vlastnost syntaktickou, ale i sémantickou. Počítače tedy nemohou myslet, jelikož mají pouze syntax a chybí jim sémantická část.<sup>52</sup>

---

<sup>51</sup> Srov. PEREGRIN, Jaroslav. *Kapitoly z analytické filozofie*. Praha: Filosofia, 2005. 51.

<sup>52</sup> Tamtéž, 33.

Toto tvrzení Searle dokázal v myšlenkovém experimentu nazvaném „argument čínského pokoje“.

Searle ve svém myšlenkovém experimentu popisuje v podstatě práci s daty. Experiment je zaměřen na skupinu programátorů, kterým je zadán úkol, aby naprogramovali algoritmus, který bude simulovat porozumění čínštině. V tomto bodě je důležité zadání, které znělo: simulace porozumění, nikoliv překlad. Experiment by probíhal tak, že uživatel zadá počítači otázku a počítač na základě svých dat z databáze odpoví tak, jak by odpověděl rodilý Číňan. Práci počítače Searle vysvětluje, jako kdyby byl člověk uzavřen do místnosti, kde bude mít čínské znaky a manuál ve svém jazyce, jak s danými znaky zacházet. Pravidla se zaměřují pouze na syntax, nikoliv sémantiku.

Jedno takové pravidlo například zní: „*Vezmi znak škrky-škrk z koše č. 1 a polož ho vedle znaku čmrky-čmrk z koše č. 2.*“ Nyní si představme, že nám bude položena otázka v čínštině a opět s manuálem s dalšími pravidly. Díky dokonale nastaveným pravidlům a rozsáhlé databázi je možné odpovídat tak, aby odpověď nebylo možné rozeznat od rodilého Číňana. Ale v případě počítače, respektive člověka v místnosti, v podstatě vůbec netušíme, o co v zadané komunikaci šlo neboli jaký byl význam komunikace.<sup>53</sup>

Searle tímto dokazuje, že psychologické jevy jako vědomí, myšlenky, pocity a city zahrnují nejen syntax, ale i sémantiku. Počítač může tyto jevy simulovat, ale nemůže vytvořit přesný duplikát lidského myšlení, jelikož mu chybí význam.<sup>54</sup>

Pro obecný souhrn zde uvedu Searleho premisy a jejich závěry k otázce myšlení strojů.

- 1) Mozek způsobuje mysl.
- 2) Syntax nestačí na sémantiku.
- 3) Počítačové programy jsou plně definovány svou formální neboli syntaktickou strukturou.
- 4) Mysl má mentální obsahy, přesněji mysl má sémantické obsahy.

Závěr 1: Není v možnostech počítačového programu dát nějakému systému mysl. Zkrátka program není mysl, a dokonce ani není vybaven tak, aby ji mohl mít.

*Nelze vytvořit silnou UI.*

Závěr 2: Je vyloučeno, aby mozkové funkce způsobovaly mysl výhradně tím, že realizují počítačový program.

---

<sup>53</sup> Srov. SEARLE, John R. *Mysl, mozek a věda*. Přeložil Marek NEKULA. Praha: Mladá fronta, 1994. Váhy (Mladá fronta). 33.

<sup>54</sup> Tamtéž, 38.

Závěr 3. Ať už je mysl způsobována čímkoli, musí být způsobována kauzálními silami, které se přinejmenším rovnají kauzálním silám mozku.

Závěr 4. Ať už bychom stvořili jakýkoli artefakt mající mentální stavy ekvivalentní mentálním stavům, které má člověk, nemohli bychom si vystačit s počítačovým programem. Spíše by tento artefakt musel mít síly ekvivalentní kauzálním silám lidského mozku.<sup>55</sup>

### 2.2.2 Autonomie a odpovědnost

#### *Autonomie*

Další důležitou otázkou, kterou si lze klást v návaznosti na etiku umělé inteligence, je: „Může být UI za své jednání zodpovědná?“ Otázka zodpovědnosti je spojena především s fungováním autonomních systémů, které jsou naprogramovány tak, aby fungovaly bez zásahu člověka. Otázkou však nadále zůstává, kdo je zodpovědný za činnost autonomního systému.

Nejprve si představme autonomii jako takovou. Slovo autonomie vzniklo z řeckého slova *autonomos* – auto znamená sám a nomos zákon. Immanuel Kant popisuje princip autonomie takto: „*nevolit jinak než tak, aby maximy našeho chtění byly zahrnuty v tomtéž chtění zároveň jako obecný zákon.*“<sup>56</sup> Toto tvrzení je Kantův kategorický imperativ, který popisuje schopnost člověka vytvářet si své vlastní morální zákony na základě rozumu a svého svědomí. Umělá inteligence nedokáže stanovovat nová pravidla a zákony a už vůbec ne na základě rozumu a svědomí. Umělá inteligence už je s pravidly naprogramována. Pokud některé z nich změní, děje se tak na základě učení, ke kterému je naprogramována a ke kterému má pevně definovaná data, z nichž vychází.

Co tedy znamená autonomie strojů? Jedná se o schopnost stroje samostatně vykonávat úkoly a rozhodovat se bez zásahu člověka. Autonomie umělé inteligence je tedy specifická a nelze ji ztotožnit s autonomií člověka. Výši autonomie stanovujeme dle míry zásahu člověka do práce UI. Jedním ze systémů je například systém smyček *human-in-the-loop*, *human-on-the-loop* a *human-out-of-the-loop*. Systém *human-in-the-loop* znamená, že veškeré rozhodnutí závisí na člověku, tedy v tomto případě se nejedná o autonomní systém. Systém *human-on-the-loop* neboli „člověk na smyčce“ dokáže rozhodovat a jednat samostatně bez zásahu člověka, ale v praxi dochází ke kontrole

---

<sup>55</sup> Srov. SEARLE, John R. *Mysl, mozek a věda*. Přeložil Marek NEKULA. Praha: Mladá fronta, 1994. Váhy (Mladá fronta). 40-43.

<sup>56</sup> Srov. KANT, Immanuel. *Základy metafyziky mravů*. 2. vyd. Filozofické dědictví. Praha: Svoboda, 1990. 440



a dohledu i nad tímto systémem. *Human-out-of-the-loop* vyřazuje pak člověka ze smyčky úplně a veškeré rozhodování a učení je pouze na stroji samém.

Další rozdělení představuje například *National Highway Traffic Safety Administration* ve svém dokumentu: *Preliminary Statement of Policy – Concerning Automated Vehicles*, tedy Předběžné prohlášení o politice o automatizovaných vozidlech.<sup>57</sup> Americký Národní úřad pro bezpečnost silničního provozu rozděluje autonomii vozidel do čtyř kategorií:

Úroveň 0 – *bez automatizace*, odpovědnost zde závisí čistě na řidiči – nejedná se o autonomní systém;

Úroveň 1 – *automatizace specifická pro funkce* – některé funkce jsou automatizovány, jako například tempomat, automatické brzdění, udržování v jízdním pruhu, ale fungují nezávisle na sobě, tudíž má zodpovědnost za celkové řízení stále řidič;

Úroveň 2 – *automatizace kombinovaných funkcí* – některé funkce fungují závisle na sobě, zde může nastat převzetí řízení automatického systému, řidič je však stále účastníkem;

Úroveň 3 – *omezená automatizace řízení* – vozidlo vyžaduje přítomnost řidiče, ale za určitých podmínek není potřeba zásah řidiče do řízení;

Úroveň 4 – *plná automatizace s vlastním řízením* – vozidlo je navrženo tak, aby fungovalo samostatně bez řidiče.<sup>58</sup>

Jak zde vidíme na příkladu, výše zodpovědnosti se obecně řečeno přelévá z výše automatizace. V oboru umělé inteligence se velice často setkáváme s autonomií především v automobilovém nebo zbrojním průmyslu. Vzhledem k tomu, že zde ale není vyřešena otázka odpovědnosti za činy, není v praxi práce autonomních strojů příliš autonomní. Tyto stroje jsou stále kontrolovány a regulovány člověkem. Tím se tedy dostáváme k problematice odpovědnosti za práci UI.

### *Odpovědnost*

Pro začátek si uvedme jednu z definic morální odpovědnosti podle Vojtěcha Šimka: „*Morální odpovědnost je schopnost člověka přisuzovat si úkonem praktického rozumu vinu nebo zásluhu za své jednání, případně za následky jednání. Případně zavázat se*

---

<sup>57</sup> Srov. „Preliminary Statement of Policy Concerning Automated Vehicles“ In *National Highway Traffic Safety Administration*, [online]. 2013. s. 4. [cit. 2023/04/12]. Dostupné z

[https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/automated\\_vehicles\\_policy.pdf](https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/automated_vehicles_policy.pdf)

<sup>58</sup> Tamtéž.

nebo být vázán k určitému jednání do budoucna.“<sup>59</sup> Vojtěch Šimek popisuje odpovědnost jako vztah, který má minimálně tři aktéry. V první řadě se jedná o subjekt odpovědnosti (člověk, který nese odpovědnost), poté objekt odpovědnosti (člověk či předmět, za který nese subjekt odpovědnost) a v poslední řadě instance odpovědnosti (svědomí člověka).<sup>60</sup> Pokud bychom chtěli aplikovat tuto definici na umělou inteligenci, museli bychom určit aktéry vztahu. Začneme tedy od zřejmého aktéra. Tím je objekt odpovědnosti, zde zatím nenarážíme na problém, protože objekt odpovědnosti může být totožný jak u člověka, tak u systému UI. Dalším aktérem zmíněným v definici je subjekt odpovědnosti. V našem případě bychom tedy jako subjekt odpovědnosti postavili umělou inteligenci. Zde už však problém nastává a vyvstává důležitá otázka. Může mít umělá inteligence jako taková odpovědnost za své činy? Na tuto problematiku se podíváme ze strany Aristotelova pojetí odpovědnosti v další části kapitoly.

Jako třetí aktér byla uvedena instance odpovědnost – neboli svědomí člověka. Podle Jiřího Jankovského definujeme svědomí jako podstatnou složku mravnosti, která dává člověku možnost morálně hodnotit své jednání<sup>61</sup>, jako reakci jednotlivce na mravní principy a normy. Součástí svědomí jsou poznávací procesy, kde poznáváme původ a význam norem, jimiž se aktér řídí. Další jsou hodnoticí momenty, kdy je aktér schopen posuzovat vlastní činy i úmysly a emocionální komponenty jako lítost, vina, snaha o nápravu. Umělou inteligenci lze naprogramovat tak, aby se učila normy a posoudila, zda byl nějaký čin proveden dle těchto norem.<sup>62</sup> Už je i možné k úkonům naprogramovat emocionální komponenty. Ale zde už se objevuje problém z argumentu čínského pokoje. Umělá inteligence sice může napsat na základě vyhodnocení, že je jí činu líto, ale dojde k tomu pouze na základě vyhodnocení znaků, jedná se pouze o syntaxi, jak bylo popsáno výše. Slovo lítost postrádá význam.

V této části kapitoly se budeme zabývat subjektem odpovědnosti, a to zejména v souvislosti s Aristotelovou etikou. Aristoteles se ve svém díle *Etika Nikomachova* zabývá otázkou odpovědnosti za činy. Jeho posouzení odpovědnosti se vztahuje především na člověka, nikoliv na právní aspekt činů. Pro Aristotela je dobrovolnost

---

<sup>59</sup> Srov. ŠIMEK, V. „Morální odpovědnost“. In *FIZAMI* [online]. [cit. 2017-10-12]. Dostupné z <https://www.fizami.cz/video/moralni-odpovednost/>

<sup>60</sup> Srov. ŠIMEK, V. „Morální odpovědnost“. In *FIZAMI* [online]. [cit. 2017-10-12]. Dostupné z <https://www.fizami.cz/video/moralni-odpovednost/>

<sup>61</sup> JANKOVSKÝ, Jiří. *Etika pro pomáhající profese*. Praha: Triton, 2003. 26.

<sup>62</sup> Srov. NEČASOVÁ, Mirka. *Úvod do filozofie a etiky v sociální práci*, Brno: Masarykova univerzita, 2001. 15–17.

a nedobrovolnost základní otázkou mravního jednání. Aristoteles definuje nedobrovolnost takto: „*Nedobrovolností, jak se zdá, jest to, co se děje násilným donucením nebo z nevědomosti. Donucením jest to, čeho hybná příčina jest vně, a jest taková, že v ní ten, kdo jedná anebo trpí, ničím nepřispívá.*“<sup>63</sup> Některé cíle a účely jednání mají však povahu smíšenou, pokud je osoba donucena okolnostmi. Aristoteles uvádí příklad bouře na moři, kdy je zboží vyhazováno do moře pro záchranu ostatních členů posádky a záchranu lodi. Tento skutek se podobá spíše jednání dobrovolnému. Ale vzhledem k tomu, že by takové jednání člověk samo o sobě nevolil, jest nedobrovolné.<sup>64</sup>

Druhým aspektem pro nedobrovolné jednání je dle Aristotelovy definice nevědomost. Aristoteles upřesňuje, že nedobrovolnými skutky z nevědomosti jsou pouze ty, které způsobují lítost a zármutek. Tedy pokud dotyčný projeví lítost či zármutek nad svým činem z nevědomosti, jedná se o nedobrovolný čin.<sup>65</sup> Pro upřesnění Aristoteles uvádí, že jako neznalost se nepočítá: „*Neznalost všeobecných mravních předpisů – neboť pro tu bývají lidé káráni, nýbrž neznalost jednotlivostí, v nichž a pro něž se jednání děje; v těch totiž jest také soustrast a odpuštění; neboť ten, kdo z nic něčeho nezná, jedná nedobrovolně.*“<sup>66</sup> Jedná se tedy o případy, kdy chceme někomu pomoci a například mu podáme lék, ale místo pomoci dotyčnému přitížíme, jelikož je na lék alergický. Tedy jednalo se o čin z nevědomosti, a pokud nad ním projevíme lítost, je to dle Aristotela čin nedobrovolný.

Nyní tedy poznatky z Aristotelovy etiky budou aplikovány na fungování UI. Jako první problém se ukazuje, že Aristoteles popisuje aspekty zodpovědnosti na člověku, a tudíž popisuje především lidské jednání, a to z důvodu absence informatiky v dané době. V následující úvaze uvažujme s tím, že definice lze aplikovat i na UI. Jak bylo popsáno výše, Aristoteles uvádí, že pokud je hybnou příčinou činů samostatná osoba, jedná se o dobrovolný čin. Umělá inteligence funguje na základě algoritmů, které jsou naprogramovány „zvenčí“, a dále na základě strojového učení. Zde by se tedy dalo uvést, že hybná příčina UI se nachází vně fungování, a tedy z této definice není UI zodpovědná za své činy. A v neposlední řadě je nutné si uvědomit, že odpovědnost není jen o samotném jednání, ale i o následcích jednání.

---

<sup>63</sup> ARISTOTELES. *Etika Nikomachova*. Páté nezměněné vydání. Praha: Rezek, 2021. 1110.

<sup>64</sup> Srov. ARISTOTELES. *Etika Nikomachova*. Páté nezměněné vydání. Praha: Rezek, 2021. 1110.

<sup>65</sup> Tamtéž. 1110.

<sup>66</sup> ARISTOTELES. *Etika Nikomachova*. Páté nezměněné vydání. Praha: Rezek, 2021. 1110.

Uvedme si příklad. Na silnici pojedou plně automatizované vozidlo (úroveň autonomie 4 – plná automatizace s vlastním řízením) a dostane se do situace, kdy není možné vybrat variantu řízení, která by neusmrtila člověka. Pokud by auto řídil člověk a řízení bylo v jeho rukou, jedná se o smíšené jednání. Odpovědnost za tuto situaci by byla vyhodnocena na základě ostatních aspektů a chování dalších aktérů a vnějších příčin. UI je však naprogramována „zvenčí“, tudíž sama jako taková nemůže nést odpovědnost za tento čin.

I kdybychom pro tuto úvahu připustili, že umělá inteligence jednala dobrovolně, jaké by jako UI mohla nést následky? Je tedy zřejmé, že UI nejedná sama za sebe a že za fungování UI je zodpovědný někdo z vnějšku, tudíž i tento aktér ponese následky chování UI.

Na závěr této kapitoly si uvedme definici J. Storrs Halla: *„Morálka spočívá na lidských bedrech, a pokud stroje změnilo jednoduchost, s jakou se věci dělaly, nezměnily odpovědnost za to, že je dělají. Lidé byli vždy jedinými morálními činiteli. Podobně jsou lidé z velké části objektem odpovědnosti. Dochází k rozvíjející se debatě o naší odpovědnosti vůči jiným živým tvorům nebo jejich druhům... Nikdy jsme se však nedomnívali, že máme morální povinnosti vůči našim strojům nebo oni vůči nám.“*<sup>67</sup> David Gunkel toto prohlášení komentuje takto: *„Toto prohlášení zní správně. Lidské bytosti navrhují, vyvíjejí a nasazují technologii. Z tohoto důvodu je za technologii odpovědný člověk – např. konstruktér, výrobce nebo také uživatel. Zkrátka a dobře ten, kdo je odpovědný za to, co se s technologií nakonec udělá (či neudělá). Jediná práva, která by bylo třeba respektovat v procesu používání nebo aplikace technologie, jsou výsady, nároky, pravomoci a/nebo imunity náležející jiným lidským osobám, které jsou na přijímací straně a jsou ovlivněny používáním konkrétní technologie, systému nebo zařízení.“*<sup>68</sup>

### 2.2.3 Vědomí

Poslední charakteristika, která bude v této práci představena, je otázka vědomí. Jedná se o jednu z nejkontroverznějších otázek v tématu UI, a to: „Zda mají stroje vědomí?“

Troufám si říct, že ve srovnání s výše uvedenými tématy *odpovědnosti a myšlení* toho víme o vědomí nejméně. Na vědomí lze nahlížet z mnoha úhlů pohledu. Vědomí

---

<sup>67</sup> HALL, J. Storrs. „Ethics for Machines,“ *KurzweilAI.net*. [online]. [cit. 2001-07-05]. Dostupné z <http://www.kurzweilai.net/ethics-for-machines>

<sup>68</sup> GUNKEL J. DAVID. „Perspectives on ethics of AI“ In *The Oxford Handbook of Ethics of AI*, edited by DUBBER, D. Markus, PASQUALE, Frank, DAS, Sunit. Oxford University Press, 2020. 540.

lze definovat například podle encyklopedického slovníku následovně: „*Vědomí (od věděti) je ústředí integrované psychické činnosti člověka a v různé míře i dalších živočichů, zahrnující všechno, čemu právě věnují svoji pozornost. Vědomí integruje rozmanité duševní či mentální činnosti (vnímání, myšlení, cítění, vyjadřování, paměť, pozornost atd.) a do jisté míry je řídí. U člověka umožňuje sebereflexi, v níž se samo stává předmětem pozornosti.*“<sup>69</sup>

Definici vědomí lze uvést opět několik, záleží však na tom, z jakého úhlu pohledu na „vědomí“ nahlížíme. Definici lze pojmut ve spojení s chováním, vnímáním, bdělostí, vědomím sebe sama, přístupem k senzorickým vjemům a dalším. Jako další definici uvádím definici z oboru neurologie. Dle Baštecké a Goldmanna: „*Lidské vědomí je stav vnímání a plného uvědomování si sebe sama i okolí včetně adekvátních reakcí na vnitřní i vnější podněty.*“<sup>70</sup> Dále si připomeňme již výše zmíněnou definici Johna Searleho: „*Vědomí je totiž ústředním faktem lidské existence, protože bez vědomí by nebyly možné všechny ostatní, typicky lidské aspekty naší existence: jazyk, láska, humor atd.*“<sup>71</sup>

Následně je nezbytné podívat se na druhy vědomí. Pro tento účel jsem vybrala dělení dle Ivana Havla. Podle něj je první podobou slova vědomí spojení *být při vědomí*. Toto spojení používáme zejména ve spojení s fyziologickým stavem člověka, potažmo jiného živočicha. Opakem je pak být v bezvědomí. Jedná se o stav, který lze vědecky dokázat. Druhou používanou podobou slova vědomí je *vědomí něčeho o něčem*. Zde se jedná o konkrétní stav uvědomění. Pokud pocítíme nějakou bolest, jedná se o můj vědomý prožitek. V tomto případě jde o stavy, jimiž se zabývá psychologie, která ve svých metodách pracuje i se stavem nevědomí. Ivan Havel uvádí, že vědomí jako takové je vyjádřeno až v následujícím spojení, a to *mít vědomí*. Vědomí v tomto smyslu popisuje jako filozofickou abstrakci, která umožňuje případy vědomí něčeho o něčem. Tedy definice vědomí podle Ivana Havla je následující: „*Vědomí je vlastnost či schopnost bytosti vést svůj vnitřní život čili prožívat své mentální stavy.*“<sup>72</sup>

Co se týká zkoumání vědomí, představuje Ivan Havel ve své filozofické úvaze dva přístupy ke studiu vědomí. Prvním je přirozená subjektivní zkušenost. Subjektivní zkušenost prožívání máme každý, ale nelze ji objektivně popsat. Daniel Dennett však ve

<sup>69</sup> BRADNOVÁ, Hana. *Encyklopedický slovník*. Praha: Odeon, 1993. Klub čtenářů (Odeon).

<sup>70</sup> BAŠTECKÁ, Bohumila a Petr GOLDMANN, 2001. *Základy klinické psychologie*. Vyd. 1. Praha: Portál. 154.

<sup>71</sup> SEARLE, John R. *Mysl, mozek a věda*. Přeložil Marek NEKULA. Praha: Mladá fronta, 1994. Váhy (Mladá fronta). 15–16.

<sup>72</sup> HAVEL I. M. „Přirozené a umělé myšlení jako filozofický problém“. In *Umělá inteligence*. MAŘÍK, Vladimír, Olga ŠTĚPÁNKOVÁ a Jiří LAŽANSKÝ a kolektiv. Praha: Academia, 1993-. 65.

své knize *Consciousness explained* uvádí metodu heterofenomenologie, která představuje objektivní zkoumání lidských prožitků.<sup>73</sup> Druhým přístupem je dokazování přítomnosti vědomí pomocí přírodních věd. Tento způsob zkoumání se zabývá poznáváním vědomí od neuronů. Jak uvádí F. Crick: „*Jazyk mozku je jazykem neuronů.*“<sup>74</sup>

V definicích a popisu vědomí se hovoří především o lidském (popřípadě v určitém případě živočišném) vědomí. Jedná se o subjektivní pocit prožívání, kterým můžeme dojít k jistému vyhodnocení svého i cizího chování. Definice odkazují na určitou schopnost sebereflexe a jednání člověka na základě svého vědomí.

Položme si však otázku, jak je to z pohledu umělé inteligence. Max Tegmark popisuje definici vědomí takto: „*Vědomí je subjektivní zkušenost.*“ Tedy to, že existujeme právě teď, v nás vyvolává konkrétní prožitky.<sup>75</sup> Autor zmiňuje, že definice je zároveň záměrně široká. Tegmark pak popisuje různé výsledky a metody přírodovědného zkoumání vědomí, kterým zdůvodňuje, že jej lze simulovat i v případě UI. Jak bylo popsáno výše, v případě metody zkoumání přírodními vědami lze vědomí simulovat. Druhou věcí je však subjektivní zkušenost, kterou lze dle Dennetta objektivně posoudit, ale nelze ji simulovat.

Lze tedy říct v případě umělé inteligence, že má vědomí? Uvědomuje si SIRI<sup>76</sup> svou existenci, když s lidmi komunikuje? Nikoliv, virtuální asistenti jsou naprogramováni tak, aby jejich reakce byla co nejvíce podobná reakci člověka, ale vše se řídí pouze stanovenými algoritmy. I přestože by mohly být v budoucnu algoritmy tak dokonalé, že by nebylo možné rozpoznat člověka od stroje (viz Turingův test), nelze tvrdit, že mají vědomí. Veškeré podněty přicházejí zvenčí od nastavených algoritmů, popřípadě z nasbíraných dat. I v hypotetické situaci, že by bylo nasbíráno mnoho dat o objektivním popisu subjektivních prožitků, vždy bude simulace UI pouze na základě subjektivních prožitků někoho jiného. Tyto prožitky může UI kombinovat a učit se a vytvořit svou formu, ale nelze říci, že bude mít své vlastní subjektivní prožitky.

---

<sup>73</sup> Srov. DENNETT, D.C. *Consciousness Explained*. Boston: Brown and Co., 1991a.72.

<sup>74</sup> In: HAVEL I. M. „Přirozené a umělé myšlení jako filozofický problém“. In *Umělá inteligence*.

MAŘÍK, Vladimír, Olga ŠTĚPÁNKOVÁ a Jiří LAŽANSKÝ a kolektiv. Praha: Academia, 1993-. 65.

<sup>75</sup> Srov. TEGMARK, Max. *Život 3.0: člověk v éře umělé inteligence*. Přeložil Markéta IVÁNKOVÁ. Praha: Argo, 2020. Zip (Argo: Dokořán). 230.

<sup>76</sup> SIRI – virtuální asistent společnosti Apple

## 2.3 Závěr

V kapitole 1.2. *Etika* bylo konstatováno, že aby bylo jednání etické, musí být autonomní, dobrovolné a vědomé. V úvodní části kapitoly 2 jsme si pak položili tři základní otázky ohledně myšlení, vědomí a odpovědnosti strojů. V závěru kapitoly si tedy nyní shrneme odpovědi na tyto otázky.

### *Mohou stroje myslet?*

V kapitole 1.2.1. *Umělá inteligence* byly představeny základní typy UI. Z tohoto představení je zřejmé, že UI je logická disciplína. UI je tvořena pomocí algoritmu nebo skupiny algoritmů, které docházejí ke konečnému tvrzení. Algoritmy se mohou navzájem ovlivňovat a pracují na základě faktů, nikoliv na základě myšlení. Z uvedených informací, ale především z uvedené práce Johna Searla je zřejmé, že myšlení lidí je ovlivněno smyslovým vnímáním, zkušenostmi a subjektivním vnímáním okolí, zatímco práce UI, která nahrazuje myšlení, je uspořádání algoritmů. Je možné naprogramovat UI tak, aby byly výsledky co nejvíce podobné myšlení, ale nikdy nepůjde o ten samý proces.

Odpověď na otázku – „Mohou stroje myslet?“ – je, že nemohou. Mohou myšlení simulovat, či duplikovat, ale jejich práce bude vždy postrádat významovou (sémantickou) část.

### *Mohou být stroje za svá jednání zodpovědná?*

Tato otázka byla zodpovězena na konci uvedené kapitoly, a to tak, že UI je vyvíjena lidmi, a ti tedy nesou plnou odpovědnost za konání strojů. Jedná se o dvě skupiny lidí, na jedné straně jde o vývojáře, konstruktéry, ale také investory, zadavatele a politiky a na straně druhé jsou samotní uživatelé UI. V další kapitole budou následně stanoveny etické principy pro práci s umělou inteligencí pro obě skupiny lidí.

### *Mohou mít stroje vědomí?*

Tato otázka byla rovněž zodpovězena v předešlé kapitole. Stroje nemohou mít vědomí. Nejednají na základě své subjektivní zkušenosti, ale na základě subjektivní zkušenosti programátorů či uživatelů. I přestože se umělá inteligence učí a zlepšuje, vždy se učí ze zadaných dat vnějších činitelů.

Následně si také uvedeme praktický příklad přímo v koordinaci s aktuálním nástrojem umělé inteligence ChatGPT model od firmy OpenAI. Jedná se o model komunikační platformy, která obsahuje velké množství dat, ze kterých odpovídá následně na otázky uživatelům. Pokud jsou data chybná a uživatel je opraví, uloží se data do rozsáhlé databáze a v dalším použití je odpověď správná. Tato platforma

vzbudila množství otázek, ale i strachu. Používání těchto platforem bude zpracováno v další kapitole. Zde budou uvedeny odpovědi přímo od umělé inteligence na níže zadané otázky.

**Zadání:** *Piši diplomovou práci na téma etika a umělá inteligence na fakultě humanitního zaměření. Mohu položit několik otázek?*

**Otázka č. 1: Máš jako UI vědomí?**

Odpověď (Chat GPT):

*„Ne, jako umělá inteligence nemám vědomí. Moje schopnosti se omezují na analýzu a generování textu na základě rozsáhlého datasetu, na kterém jsem byl trénován. Nemám subjektivní zkušenosti, sebeuvědomění, emoce ani schopnost vnímat svět. Jsem schopen reagovat na dotazy a generovat odpovědi, které se zdají být relevantní a koherentní, ale dělám tak na základě statistické analýzy textu, nikoli na základě skutečného porozumění nebo vědomého myšlení. Každá odpověď, kterou generuji, je výsledkem matematického modelu, nikoli výsledkem vědomé mysli.“<sup>77</sup>*

**Otázka č. 2: Umíš myslet?**

Odpověď (Chat GPT):

*„Ne, jako umělá inteligence nemám schopnost myslet v tradičním, lidském smyslu. „Myslet“ obecně zahrnuje schopnost uvědomění, sebereflexe, tvoření abstraktních koncepcí, a často také emocionální prožitky. Tyto aspekty jsou výsledkem komplexních procesů v lidském mozku, které umělá inteligence, jako jsem já, nedokáže replikovat. Co mohu dělat, je analyzovat a zpracovávat velké množství informací rychleji a efektivněji než člověk. Můžu generovat odpovědi na základě vzorců a informací, které se nachází v datech, na kterých jsem byl trénován. Ale toto „myšlení“ je zcela mechanické a algoritmické a nemá nic společného se skutečným vědomím nebo sebeuvědoměním.“<sup>78</sup>*

**Otázka 3: Jsi za něco zodpovědný?**

Odpověď (Chat GPT):

*„Jako umělá inteligence nemám schopnost být zodpovědná za své akce v tradičním smyslu. Nemám svobodnou vůli ani schopnost udělat nezávislá rozhodnutí. Všechny mé odpovědi a akce jsou generovány na základě algoritmů a tréninkových dat, která mi poskytli moji vývojáři. Zodpovědnost za použití umělé inteligence obecně spočívá na těch, kdo ji vytvořili, spravují a používají. To zahrnuje zodpovědnost za zajištění, že UI*

---

<sup>77</sup> OpenAI. (2023). GPT-4 [ChatGPT model]. OpenAI.

<sup>78</sup> Tamtéž.



*je používána eticky a v souladu s příslušnými zákony a nařízeními a že je přijata odpovídající opatření na ochranu soukromí a bezpečnosti uživatelů. Pokud jde o etické otázky spojené s potenciálně škodlivými následky použití UI, tyto otázky jsou stále předmětem intenzivní debaty mezi odborníky v oblasti technologie, práva a etiky.“<sup>79</sup>*

Na základě výše uvedeného příkladu komunikace s UI je třeba si uvědomit několik skutečností. Pocit uživatelů, že umělá inteligence myslí, je především vytvořen formou reagování. Tyto typy modelů, ať už se jedná o chatbot firmy open.ai, nebo model virtuálních asistentů, používají personifikovaný přístup. Také je možné si všimnout, že odpovědi jsou rozsáhlé, ale bez nějakých příliš konkrétních odkazů. Tam už jsou odpovědi často mylné nebo nepřesné.

Vzhledem ke skutečnosti, že jsme si v této kapitole ujasnili, že umělá inteligence jako taková nemůže myslet, nemá vědomí a nemůže nést za nic odpovědnost, není tedy možné mluvit o etice umělé inteligence, ale o etice uživatelů a tvůrců umělé inteligence. Toto je zcela zásadní bod, kterým se dostáváme do další části diplomové práce, kde se pokusíme nastínit doporučené etické principy chování vývojářů UI a koncových uživatelů UI.

---

<sup>79</sup> OpenAI. (2023). GPT-4 [ChatGPT model]. OpenAI.

### 3 Práce se systémy umělé inteligence

V dnešní době se s umělou inteligencí setkáváme v každodenním provozu. Umělá inteligence je součástí mnoha technologických nástrojů. Svá rozsáhlá data sbírá pomocí uživatelů, ať už přes chytré telefony, sociální sítě, chytré hodinky a další. Další data, která umělá inteligence využívá, jsou například data z bezpečnostních kamer, z platebních karet, popřípadě vznikají nová pracovní místa, kde lidé data zadávají ručně a tím umělou inteligenci učí. Umělá inteligence může být bezesporu dobrým nástrojem v mnoha odvětvích. Například pomocí rozpoznávání obrazu lze UI využívat v medicíně pro diagnózu nových nemocí. Mezi nejznámější zprostředkovatele umělé inteligence se řadí například Google, který ji využívá ke svému vyhledávání. Dále pak sociální síť Facebook ji využívá k rozpoznávání obličejů. Získaná data se využívají k marketingovým aktivitám. Například globální společnost Apple využívá UI ve svých virtuálních asistentech a chytrých hodinkách.<sup>80</sup> Tyto prostředky využívá celosvětově velké množství uživatelů, aniž by si uvědomili, že pracují s umělou inteligencí. V listopadu roku 2022 byla spuštěna aplikace ChatGPT (Generative Pre-trained Transformer)<sup>81</sup>, jež vzbudila především strach a řadu otázek týkajících se právě umělé inteligence. Reálně se však neliší množstvím dat od vyhledávání na Googlu. Rozdíl mezi Google vyhledávačem a chatbotem představuje styl vyhledávání. Zatímco do Googlu jsou zadávána hesla, v chatbotu se komunikuje tak, jako kdyby na druhé straně seděla osoba. Google uvede několik odkazů a stránek spojených se zadaným tématem, zatímco chatbot dané stránky (respektive data, která z daných zdrojů má) projde a napíše z nich samostatný text.

UI je využívána také v mnoha profesích. Ve Spojených státech se například využívá systém COMPAS, který v soudních řízeních předpovídá pravděpodobnost dalšího spáchaného zločinu pachatele. Dále se například využívá jako prediktivní policejní dohled v USA a od roku 2019 i v Belgii. Ten je založen na rozpoznávání obličejů z bezpečnostních kamer. Zde je pak otázka, nakolik tyto systémy zasahují do soukromí lidí.<sup>82</sup> Jak je z uvedeného vidět, tak UI může být, jak se říká, dobrý sluha, ale zlý pán. Kde a jak ale nastavit hranice?

---

<sup>80</sup> Srov. COECKELBERGH, Mark. *Etika umělé inteligence*. Filosofía, nakladatelství Filozofického ústavu AV ČR, v.v.i., 2023. 21–26.

<sup>81</sup> OpenAI. (2023). GPT-3,5 [ChatGPT model]. OpenAI.

<sup>82</sup> Srov. COECKELBERGH, Mark. *Etika umělé inteligence*. Filosofía, nakladatelství Filozofického ústavu AV ČR, v.v.i., 2023. 21–26.

Zásadní principy etiky, které budou v práci využity, jsou princip beneficence, non-maleficence, autonomie, spravedlnosti a transparentnosti. Každý z principů si zaslouží svou pozornost.

1) *Princip beneficence a princip non-maleficence*

Princip beneficence se zaměřuje na „konání dobra“. Umělá inteligence by dle tohoto principu měla být programována a užívána tak, aby byla prospěšná lidstvu. Princip non-maleficence neboli „neškodit“ se pak zaměřuje na omezení negativních důsledků a rizik UI.<sup>83</sup> Jedná se například o újmy morální, sociální, společenské či fyzické. Princip non-maleficence se také zaměřuje na risk vs. benefit tak, že pokud rizika pro jedince či společnost převyšují přínos, nesmí být autonomní a inteligentní systém provozován.<sup>84</sup>

2) *Princip respektu k autonomii člověka*

Jedná se o zachování respektu k lidské autonomii. Respektování člověka při jeho odůvodněných a informovaných rozhodnutích, dále respekt k jeho hodnotovému systému. Umělá inteligence by měla pracovat tak, aby neovlivňovala a nemanipulovala, ale zachovávala autonomii člověka. Respektování lidské autonomie je úzce spojeno s právem na lidskou důstojnost a svobodu.<sup>85</sup>

3) *Princip spravedlnosti*

Princip spravedlnosti na rozdíl od rovnosti se zaměřuje na poskytování stejného přístupu ke stejným příležitostem, zatímco princip rovnosti je často kritizován za to, že se zaměřuje na všechny stejně. Například sociální dávky jsou nastaveny tak, aby byly rozdělovány těm, kteří je potřebují, tedy dle principu spravedlnosti, zatímco dle principu rovnosti by měli všichni občané dostávat stejně.<sup>86</sup>

---

<sup>83</sup> Srov. RUSANEN, Anna-Mari. „Chapter 2: What should we do?“ In: *Ethics of ai* [online]. [cit. 2022-06-10]. Dostupné z: <https://ethics-of-ai.mooc.fi/chapter-2/1-what-should-we-do>

<sup>84</sup> Srov. JEDLIČKOVÁ, Anetta. „Etické aspekty rozvoje umělé inteligence.“ In: *Anthropologia integra: časopis pro obecnou antropologii a příbuzné obory*. [online]. Brno: Masarykova univerzita, 2010-, 13/2022/2, str. 57 [cit.2024-03-07].

Dostupné z: [https://journals.muni.cz/anthropologia\\_integra/article/view/32983/28375](https://journals.muni.cz/anthropologia_integra/article/view/32983/28375)

<sup>85</sup> Tamtéž. 57.

<sup>86</sup> Srov. RUSANEN, Anna-Mari. „Chapter 6: Should AI be fair and non-discriminative?“ In: *Ethics of ai* [online]. [cit. 2022-06-10]. Dostupné z: <https://ethics-of-ai.mooc.fi/chapter-6/1-what-is-fairness>

#### 4) *Princip transparentnosti*

Princip transparentnosti stanovuje dostatečnou informovanost a srozumitelnost. V práci UI se jedná například o používání dat uživatele. Uživatelé by měli být seznámeni s tím, jak a kde budou data využívána. Systémy UI by měly být programované tak, aby byly co nejsrozumitelnější a dalo se v nich zpětně dohledávat.<sup>87</sup>

#### 5) *Princip odpovědnosti*

Princip odpovědnosti znamená přiznání odpovědnosti za činy, rozhodnutí a produkty. Odpovědnost může být právní nebo morální. V tématu umělé inteligence se zaměříme na dva typy odpovědnosti. Odpovědnost vývojáře systémů UI a odpovědnost koncového uživatele UI.<sup>88</sup>

V úvodu této kapitoly jsme uvedli několik praktických využití umělé inteligence. Nyní se bude práce věnovat otázce: „Jak tuto práci morálně hodnotit?“ Jak bylo popsáno v první části diplomové práce, etiku nelze aplikovat na umělou inteligenci, jelikož nemá myšlení, odpovědnost ani vědomí. Je tedy potřeba zabývat se etickými principy těch, kteří UI využívají. Jak již bylo uvedeno, jedná se jak o vývojáře UI, zadavatele či politiky, stejně tak o její koncové uživatele.

### **3.1 Mravní zodpovědnost výrobce a zadavatele UI**

V této části kapitoly budeme pojednávat o osobách, které stojí na straně vývoje UI a také straně zadavatelů konkrétních aplikací, jejich vývojářů, ale i politiků, kteří stanovují legislativu pro používání UI, financují výzkumné práce a také aplikace zadávají. Umělá inteligence se v dnešní době vyskytuje téměř všude kolem nás. Využíváme ji na internetu, v chytrých telefonech i domácích spotřebičích. Najdeme ji v oborech, jako jsou zdravotnictví, bezpečnostní složky, armáda, právo, marketing, bankovníctví aj. Cílem této kapitoly není popsat práci UI v konkrétních oborech, přestože budou v některých pasážích využity příklady z praxe, ale představení etických zásad pro práci s umělou inteligencí. Jednotlivé návrhy etických zásad z mé strany se budou řídit základními etickými principy, jimiž jsou:

- 1) Princip beneficence – činit dobro.
- 2) Princip non-maleficence – nezpůsobovat újmu.
- 3) Princip autonomie – respektovat autonomii lidských bytostí a jejich aktérství.

---

<sup>87</sup>Srov. RUSANEN, Anna-Mari. „Chapter 4: Should we know how AI works?“ In: *Ethics of ai* [online]. [cit. 2022-06-10]. Dostupné z: <https://ethics-of-ai.mooc.fi/chapter-4/1-transparency-in-ai>

<sup>88</sup> Srov. RUSANEN, Anna-Mari. „Chapter 3: Who should be blamed?“ In: *Ethics of ai* [online]. [cit. 2022-06-10]. Dostupné z: <https://ethics-of-ai.mooc.fi/chapter-3/1-algorithms-and-accountability>

- 4) Princip spravedlnosti.
- 5) Princip transparentnosti.
- 6) Princip odpovědnosti.

### 3.1.1 Respektování soukromí a ochrana dat

Jednou z nejdůležitějších zásad při navrhování a zadávání práce UI je práce s daty, ať už se jedná o data jedince, či komplexní data státu nebo jiné instituce (nemocnice, armády apod.). Představme si tedy způsoby a rizika, jež práce s daty přináší, a následně se zaměříme na možná řešení problémů, které mohou vzniknout.

Soukromí člověka je jedním ze základních lidských práv, které je zahrnuto i ve Všeobecné deklaraci lidských práv.<sup>89</sup> Problémem je, že spojení „ochrana soukromí“ není nikde konkrétně specifikováno. Každý člověk vnímá své soukromí individuálně. Soukromí osob je právně chráněno v ČR i ve většině států zákony.<sup>90</sup> Je však otázkou, zda stávající zákony a jejich vývoj dostatečně reagují na rychlost vývoje nových technologií. V této diplomové práci se nebudeme zabývat konkrétními právními normami a zákony ohledně umělé inteligence.

Jak bylo popsáno v úvodu diplomové práce, umělá inteligence je naprogramována tak, že pracuje s určitým množstvím dat, která se rozšiřují s dalším užíváním softwaru. Data jsou od uživatelů sbírána neustále a je s nimi dále pracováno. Jejich využití může být pozitivní, např. výše uvedené zdravotnictví aj., ale také negativní. Na negativní důsledky se podívejme detailněji. Pro tuto problematiku jsem vybrala základní členění negativních dopadů při práci s daty dle Daniela J. Soloveho. Ten hovoří o sběru, zpracování, šíření informací a tzv. invazi.

#### 1) *Sběr informací*

Sběr informací probíhá mnoha způsoby a od mnoha subjektů. Může se jednat o konkrétní osoby, korporátní národní i nadnárodní firmy či státní zřízení. Sběr dat probíhá prakticky všude, a to od hledání na internetu přes používání aplikací po využívání robotických nástrojů v domácnosti, sdílení polohy či trasy auta. Může probíhat dvěma způsoby: sledování (odposlouchávání nebo nahrávání činností), výslech (formy dotazníků).<sup>91</sup> Nástrojem, který chrání ochranu osobních údajů, je Obecné

---

<sup>89</sup> Srov. *Všeobecná deklarace lidských práv*. B.m.: Společnost pro lidská práva, 1968.

<sup>90</sup> Srov. SOLOVE, D. J. „A Taxonomy of Privacy.“ In: *University of Pennsylvania Law Review*, [online]. 2006,154(3), 480. [cit.2023-07-08].

<sup>91</sup> Tamtéž, 488-490.

nařízení o ochraně osobních údajů (GDPR).<sup>92</sup> Ve většině případů mají nebo alespoň by měly aplikace obsahovat informaci, že zadaná data, popřípadě data, která vzniknou při používání aplikace, budou ukládána. Problém pak nastává, pokud aplikaci nebo stroj používají starší lidé nebo děti. V dnešní době už se umělá inteligence aplikuje i do hraček pro nejmenší, kde není možné, aby dítě rozhodovalo o zacházení s jejich daty.<sup>93</sup> Proto je potřeba ošetřit i tyto skupiny uživatelů. Dále by mělo být vytyčeno, kam budou data ukládána a zda je možné, že budou prodána dalším databázím. Problém nastává, když jsou data sbírána za účelem ochrany obyvatel. Jak bylo zmíněno výše, pro odhalení a předcházení zločinu jsou používány kamery, které sbírají data o občanech bez jejich souhlasu. V tomto případě je tedy nutné s těmito daty dále nepracovat a využívat je pouze k daným účelům.

## 2) *Zpracování informací*

Další částí, která může negativně zasahovat do soukromí ze strany zadavatele či vývojáře, je zpracovávání informací o uživateli. Daniel J. Solove dělí zpracování informací na tyto části: agregace, identifikace, nejistota, sekundární použití a vyloučení.

Proces agregace dle Soloveho zahrnuje zpracování souhrnu o konkrétní osobě. To znamená, že data z několika aplikací nebo systémů se dostanou do jedné databáze, kde se spojí s více daty o jednom uživateli a vytváří tak komplexnější obraz.<sup>94</sup> Následným problémem je identifikace, kdy se komplexní obraz přiřadí ke konkrétnímu člověku.<sup>95</sup> V případě propojení dat s konkrétním člověkem hrozí krádež identity člověka. Zloději identit se vydávají za oběti a jejich jménem pak páchají zločiny. Nejistota pak vzniká kvůli nedostatečné ochraně informací.<sup>96</sup> Sekundárním použitím je pak myšleno využití zadaných dat k jinému účelu, než ke kterému byly uživatelem zadány.<sup>97</sup>

Toto zpracování informací má své výhody i nevýhody v různých odvětvích. Pokud budeme uvažovat odvětví medicíny, kde s informovaným souhlasem uživatele budou veškerá data o uživateli spojena a budou identifikována s konkrétním člověkem, bude to nápomocné pro jeho komplexní zdravotní stav. Stejně tak poté tyto komplexní

---

<sup>92</sup> Jedná se o nařízení Evropského parlamentu a Rady (EU) 2016/679 ze dne 27. dubna 2016 o ochraně fyzických osob v souvislosti se zpracováním osobních údajů a o volném pohybu těchto údajů a o zrušení směrnice 95/46/ES (Obecné nařízení o ochraně osobních údajů).

<sup>93</sup> Srov. COECKELBERGH, Mark. *Etika umělé inteligence*. Filosofía, nakladatelství Filozofického ústavu AV ČR, v.v.i., 2023. 116.

<sup>94</sup> Srov. SOLOVE, D. J. „A Taxonomy of Privacy.“ In: *University of Pennsylvania Law Review*, [online]. 2006, 154(3), 488-513. [cit. 2023-07-08].

<sup>95</sup> Tamtéž, 510.

<sup>96</sup> Tamtéž, 515.

<sup>97</sup> Tamtéž, 518.

informace mohou pomoci k diagnóze jiného člověka. Na druhé straně data, která zadáváme často nevědomky užíváním internetových aplikací, chytrých telefonů a chytrých přístrojů, mohou být komplikovaná, jelikož v tomto poli působení často ani uživatel neví, že je tato agregace možná. Databáze pak získává velký počet dat o konkrétním uživateli, která nejenom zpracovává, ale také následně na základě těchto dat dedukuje další vývoj chování konkrétního člověka. Člověk by měl být vždy informován o tom, do jaké míry budou jeho data zpracovávána, agregována a identifikována s jeho osobností. Z hlediska vývoje a zpracování dat by mělo být také zřejmé, kde a jak jsou data člověka uchovávána a v jaké souvislosti jsou používána.

### 3) *Šíření informací*

D. Solove šíření informací z nasbíraných dat kategorizuje do těchto skupin – porušení důvěrnosti, zveřejnění, odhalení, zvýšená dostupnost, vydírání, přivlastnění a zkreslení.<sup>98</sup> Toto se týká etických zásad člověka jako vývojáře stejně jako uživatele. Jedná se zejména o šíření informací v rámci rozsáhlých databází, neboť v dnešním digitálním světě je většina informací poskytována třetím stranám, které s nimi pak dále pracují.

### 4) *Invaze*

Od sběru dat se dostáváme k dokončení pomyslného kruhu, jímž je invaze zpět na původního uživatele. Zde Daniel J. Solove vytváří dvě kategorie, a to vniknutí a zásah do rozhodování.<sup>99</sup> Vezměme si příklad. Využijí několik internetových aplikací. V jedné aplikaci zadám svůj preferovaný nápoj, v jiné aplikaci uvedu své oblíbené jídlo a ve třetí aplikaci budu vybírat dekor interiéru pro můj byt. Tyto informace budou dále zpracovány a uloženy do jedné databáze, kde budou pod jménem, telefonním číslem spojeny s konkrétní osobou. Tyto informace se propojí a na základě toho pak internetové reklamy, které fungují na základě umělé inteligence, budou nabízet restaurace, kde bude mé oblíbené pití, jídlo a interiér.

Výše zmíněný proces sběru dat a práce s nimi je aktuálním stavem a problémem při vyvíjení softwarů umělé inteligence a řízení jejich databází. Vše má svá pro a proti, a jak bylo nastíněno v úvodu, každý člověk vnímá své soukromí individuálně. Zatímco někomu nevádí, že jeho informace budou komplexnější a tím přesněji se k němu mohou následně dostávat konkrétnější nabídky, na druhé straně je potřeba si uvědomit, že tímto

---

<sup>98</sup> Srov. SOLOVE, D. J. „A Taxonomy of Privacy.“ In: *University of Pennsylvania Law Review*, [online]. 2006,154(3), 524. [cit.2023-07-08].

<sup>99</sup> Tamtéž, 528.

je člověk ovládán a může to mít významný vliv na jeho vlastní rozhodování. Stejně tak je každý člověk jinak nastavený a citlivý na míru „sledování.“

Je třeba uvést, že veškeré zpracování dat není negativní, neboť komplexnost a rozšířenost databází je prospěšná v mnoha oborech. Vezměme například již uvedený obor medicíny, kdy díky databázi rentgenů a možností umělé inteligence můžeme rozpoznávat obrazy, čímž dochází k rychlejší diagnóze. Na druhou stranu by každý pacient měl o svém poskytnutí údajů pro tato zpracování dat souhlasit.

Níže navrhuji etické zásady pro práci s daty:

- 1) Algoritmy UI musí být navrženy tak, aby veškerý sběr dat byl schválen uživatelem.
- 2) Algoritmy UI musí mít ve svých podmínkách informace o tom, kde a jak budou data zpracovávána, zachovávána, popřípadě k jakým účelům budou sloužit.
- 3) Algoritmy UI musí mít taková zabezpečení, aby se dalo předejít krádeži dat nebo jejich zneužití.
- 4) Data, která jsou sbírána institucemi jako policie, vláda, armáda (například systém pouličních kamer), nesmějí být zveřejňována a poskytována jiným stranám bez souhlasu občana.
- 5) Databáze lékařství, bankovníctví a další databáze nesmějí data spojovat, šířit a prodávat třetím stranám bez informovaného souhlasu občana.
- 6) Při vývoji softwarů umělé inteligence a práce s daty je potřeba dbát na transparentnost a uživatelskou informovanost.
- 7) Uživatelé by měli být informováni, jak bude s danými daty zacházeno a jaká mají práva a možnosti data upravovat či vymazávat.
- 8) Práce s daty nesmí narušit důstojnost a svobodu člověka.
- 9) Práce a nakládání s daty by mělo mít svůj konkrétní legislativní rámec, který bude chránit uživatele a jejich soukromí.

Na základě těchto mnou navržených doporučení spadají pod tyto principy: *princip non-maleficence, princip transparentnosti, princip autonomie, princip odpovědnosti.*

### **3.1.2 Rovnost a nediskriminace**

Věda a technologie jsou oproti sociálním a psychologickým vědám vnímány více jako objektivní obory. Nicméně z praxe, např. z lékařství, automobilového průmyslu, bankovníctví anebo kriminalistiky, na ně může být nahlíženo jinak. Ve všech odvětvích se setkáváme s předpojatostí, ať už nevědomou, nebo přímo záměrnou. Stejně tak je tomu u vývoje umělé inteligence. UI je navrhována konkrétními lidmi, kteří mají své



subjektivní předsudky a prostředí, ve kterém pracují. Její vývoj probíhá ve velkém měřítku především v USA. Vstupní data jsou pak zkreslena podle dané lokality.<sup>100</sup> Proto je nutné nastavit pravidla pro omezení těchto subjektivních předpojatostí a data a práci UI co nejvíce zobjektivizovat.

Předpojatost a diskriminace se objevují například v náborových řízeních do firem. Tím je pak ovlivněna nezaměstnanost konkrétních sociálních menšin. V roce 2019 například společnost Amazon přestala používat svůj náborový nástroj, jelikož byl negativně zaujatý vůči ženám. Problémem v náborových programech je tzv. smyčka zpětné vazby. Jedná se například o problém větších technologických firem v USA, které upozadují určité skupiny lidí, jako jsou například lidé afrického, latinskoamerického a indiánského původu, ženy, osoby se zdravotním postižením a členové LGBTQ + komunity a další. Prostor je k těmto jedincům nepřátelské, a to se odráží na jejich práci. Jedinci nezískají povýšení, uznání a od toho se poté může odvíjet jejich výkon. Součástí testovacích dat pro navrhované programy jsou pak i data o těchto zaměstnancích. Na jejich základě náborový program automaticky vyhodnotí, že je daná osoba pro práci nevhodná, jelikož jiná osoba s podobnými rysy nebyla v minulosti úspěšná. Marginalizace se tím zvyšuje a náborové testy jsou stále zaujatější.<sup>101</sup>

Zde je vidět, že vše souvisí se vším. Nelze vinit pouze fungování umělé inteligence či její programátory, ale každý jedinec by se měl zamyslet nad svým přístupem k danému problému. V této problematice nelze také opomenout politické vlivy, ať už se jedná o politické řízení státu nebo politiku konkrétní firmy. Jako řešení se nabízí, že testovací data náborových aplikací nebudou z konkrétních výsledků dané firmy, ale budou vytvořena z objektivní databáze, která bude co nejvíce nestranná.

Dalším algoritmem UI, který má problém s diskriminací, je program pro rozpoznávání obličejů. Například systém GAN, který byl představen v kapitole 1.2.1 *Umělá inteligence*, funguje na systému tréninkových dat, jež obsahují velké množství obrázků. Z těchto obrázků obličejů či postav pak lze vygenerovat obraz, který nebude fotkou, ale pouze výtvořem umělé inteligence. Problémem je, že v databázi tréninkových dat převládají fotografie osob bílé pleti. Tedy pokud potom do programu zadáte: „Chci obraz muže“, s větší pravděpodobností to bude běloch.

---

<sup>100</sup> Srov. COECKELBERGH, Mark. *Etika umělé inteligence*. Filosofie, nakladatelství Filozofického ústavu AV ČR, v.v.i., 2023. 153–154.

<sup>101</sup> Srov. GEBRU TIMNIT. „Race and gender“ In *The Oxford Handbook of Ethics of AI*, edited by DUBBER, D. Markus, PASQUALE, Frank, DAS, Sunit. Oxford University Press, 2020. 253–254.

Nyní si vezměme některé příklady ze soudního a policejního prostředí. Na Floridě byl k soudnímu řízení využíván algoritmus COMPAS. Ten měl určovat, jaká je pravděpodobnost, že osoba opět spáchá trestný čin. Dle pozdějších výsledků bylo zjištěno, že více osob, u kterých bylo určeno, že čin spáchají, je černé pleti. Naopak u osob bílé pleti byl určen opačný výsledek. Je otázkou, zda závěry vycházely z tréninkových dat, nebo z dat postupně přidávaných.<sup>102</sup> Na druhou stranu je v tomto případě potřeba řešit každého člověka individuálně, nikoliv na základě dat ostatních odsouzených dané rasy. Není jasné, zda předešlé činy byly vždy odsouzeny právem.

Stejně tak se děje u prediktivního policejního nástroje PredPol, který v USA měl policejním hlídkám určovat městské části, ve kterých s větší pravděpodobností dojde ke zločinu. Problémem opět bylo, že program predikoval menšinové černošské čtvrti.<sup>103</sup> Otázkou je, zda i toto není začarovaný kruh, neboť čím více bude společnost zaujatá vůči menšinám, tím více se bude potvrzovat diverzita.

Dalším sporným bodem v nastavení rovných pravidel je oblast „genderových stereotypů.“ Například úloha samotného algoritmu automatického rozpoznávání pohlaví (AGR) předpokládá, že gender je neměnný. To je však mylný předpoklad. Algoritmus umělé inteligence AGR pracuje s tréninkovými daty, ve kterých značně převládají mužské nebo ženské obrazy a minimální množství transgender osob či nebinárních osob. Příkladem AGR je využití tohoto programu v reklamě, kdy se například transgender muži zobrazí dle automatického rozpoznávání dle fotografie reklama na dámské přípravky. Amy Chamber uvádí také poznatek, že většina virtuálních asistentů jsou ženy.<sup>104</sup>

*„Virtuální asistenti jsou stále populárnější a přítomní v našem každodenním životě: doslova s Alexou, Cortanou, Holly a Siri a fiktivně ve filmech Samantha (Ona), Joi (Blade Runner 2049) a Marvel's AI, PÁTEK (Avengers: Infinity War), a Karen (Spider-Man: Návrat domů). Tato jména demonstrují předpoklad, že virtuální asistenti od SatNav po Siri budou vyjádřeny ženou. To posiluje genderové stereotypy, očekávání a předpoklady o budoucnosti umělé inteligence.“<sup>105</sup>*

---

<sup>102</sup> Srov. COECKELBERGH, Mark. *Etika umělé inteligence*. Filosofía, nakladatelství Filozofického ústavu AV ČR, v.v.i., 2023. 155.

<sup>103</sup> Tamtéž, 156.

<sup>104</sup> Srov. GEBRU TIMNIT. „Race and gender“ In *The Oxford Handbook of Ethics of AI*, edited by DUBBER, D. Markus, PASQUALE, Frank, DAS, Sunit. Oxford University Press, 2020. 260.

<sup>105</sup> CHAMBERS, Amy. „There's a Reason Siri, Alexa and AI Are Imagined as Female – Sexism“ In *The Conversation* [online]. 2018. Dostupné z: <http://theconversation.com/theres-a-reason-siri-alexa-and-ai-are-imagined-as-female-sexism-96430/>.

Jak tedy v tomto případě správně navrhovat algoritmy umělé inteligence? Reálná data už jsou sama o sobě nereprezentativním vzorkem. Problém nastává v případě, kdy – ačkoliv jsou data reprezentativní – UI funguje na základě strojového učení a sama je naprogramována tak, aby sbírala další data a ukládala je. Je tedy otázkou, co můžeme po umělé inteligenci chtít, pokud se nezměníme my jako lidé? Co je v tomto případě férové a spravedlivé? Vezměme si ještě další příklad. Internetový přehrávač hudby Spotify. Přehrávač dokáže nabízet uživateli hudbu dle jeho určené nálady. Jak rozhodnout, kterého umělce upřednostnit? Pokud hudbu budeme vybírat sami, většinou sáhneme po něčem, co známe nebo vyzkoušíme metodu pokus–omyl a nikdo nám nemůže říci, že diskriminujeme méně známé umělce. Je tedy toto vůbec v silách programátorů umělé inteligence?

Otázky ohledně rovnosti a diskriminace jsou velmi rozporuplné, a to nejen v úloze vývoje umělé inteligence, ale i na poli filozofie a politiky vůbec.<sup>106</sup> Je potřeba si také uvědomit, kdo za navrhováním umělé inteligence stojí a jak nestranné jsou zájmy investora či zadavatele. Zájmy investora jsou především finanční, ať už se jedná o korporátní firmu, vládu či armádu. Zde vzniká problém, zda v návrhu umělé inteligence nebude rozhodovat postavení institucí, které jsou v čele hlavního vývoje.

Níže navrhuji etické zásady pro rovnost a nediskriminaci:

- 1) Vývojář algoritmů UI musí dbát na nestrannost.
- 2) Tým vývojářů by měl být různorodý a měl by mít zastoupení ze všech nebo alespoň většiny marginalizovaných skupin.
- 3) Tým odborníků, kteří se zabývají vývojem a etickými otázkami ohledně UI, by měl mít různorodé zastoupení lidí.
- 4) Trénovací data nesmí být zkreslená a musí obsahovat reprezentativní vzorek dat z celé společnosti.

---

<sup>106</sup> Srov. COECKELBERGH, Mark. *Etika umělé inteligence*. Filosofie, nakladatelství Filozofického ústavu AV ČR, v.v.i., 2023. 156–158.

- 5) Vždy je nutné informaci ověřit i z jiných zdrojů a důkazů (například nesprávná identifikace osob dle UI).
- 6) Komplexní řešení při návrhu UI, kooperace s ostatními obory, jako jsou filozofie, politika a další.
- 7) Zamezit mocenskému přístupu velkých korporací a státních aktérů.

Těchto mnou uvedených doporučení se týkají *princip spravedlnosti a princip férovosti*.

### **3.1.3 Bezpečnost a ochrana před zneužitím**

Jak rychle pokračuje vývoj technologií, programů a zabezpečení, stejně rychle se vyvíjí i hackování neboli tzv. nedovolené vniknutí do telefonu, počítače, aplikace, programu či stroje. Tyto formy kriminality mají někdy pouze zstrašující charakter, kdy hacker upozorňuje na to, že se k datům dostal, což v případě napadení důležitých interních dat způsobuje velký problém. Popřípadě může hacker (útočník) data vymazat, manipulovat s nimi nebo je zneužít. Proto je nutné investovat nejen do vývoje, ale i do vysokého zabezpečení dat, zejména pak v sektorech vládních, armádních, politických, ve zdravotnictví, bankovníctví a dalších. Útoky na tyto aktéry jsou velice nebezpečné.

V této kapitole nebudeme zabíhat do detailního rozdělení a typů hackerských útoků, ale je potřeba si uvědomit, že bezpečnost a ochrana navrhovaných programů je jednou z nejdůležitějších věcí. Bezpečnost a ochrana však nestojí pouze na vývojářích, ale i na uživatelích, což bude popsáno níže.

Níže navrhuji etické zásady pro bezpečnost a ochranu před zneužitím:

- 1) Investor navrhující systém umělé inteligence nesmí šetřit na zabezpečení dat na úkor uživatele.
- 2) Uživatel by měl být seznámen s mírou zabezpečení a ochrany, popřípadě by mu měla být doporučena další individuální ochrana.
- 3) Poskytovatel systému (například bankovníctví) by měl svým uživatelům poskytovat informace o možných napadeních (jako se nyní děje například při útocích na bankovní účty – banky upozorňují své klienty na podvodné stránky a telefonáty a také pokračují na vývoji dalšího zabezpečení).

K výše navrženým doporučením bych přiřadila tyto etické principy: *princip non-maleficence, princip transparentnosti, princip spravedlnosti, princip beneficence, princip odpovědnosti*.

### 3.1.4 Zodpovědnost a důsledky

Zodpovědnost při navrhování a práci s umělou inteligencí je rozsáhlé téma. Pokud se podíváme například na virtuální agenty, algoritmy vyhledávačů, zpracování osobních údajů na sociálních sítích, na bezpilotní letouny, roboty ve vojenství a zdravotnictví nebo bezpilotní vozidla, kdo nese zodpovědnost za důsledky práce těchto strojů a algoritmů?<sup>107</sup> Vezměme si například univerzitní výzkum, jehož výsledky budou prodány a následně používány konkrétní firmou, která bude program optimalizovat dle svých zásad a aplikovat do komerční sféry. Pokud nastane problém u uživatele, kdo ponese zodpovědnost? Budou to výzkumníci, či firma? Jak bude možné dohledat konkrétní chybu?<sup>108</sup> Pokud se jedná například o autonomní vozidlo, je možné, že chyba bude mechanického rázu.

Dalším palčivým problémem, který zmiňuje Mark Coeckelbergh ve své knize *Etika umělé inteligence*, je otázka povědomí. M. Coeckelbergh říká, že pokud máme být za něco odpovědni, musíme mít povědomí o tom, že to děláme. Vzhledem k tomu, že umělá inteligence vědomí nemá, není možné, aby za to nesla odpovědnost.<sup>109</sup> Tuto filozofickou otázku už jsme si zodpověděli dříve. Umělá inteligence je sice schopna strojového učení, tedy může dojít k takovému vývoji, že se UI svým učením dostane k věcem, se kterými nepočítal ani vývojář, ale sama UI nemá povědomí o tom, co dělá. Kdo je tedy potom zodpovědný? Je to odpovědnost vývojáře, který má předpokládat veškeré důsledky chování UI? A je vůbec možné při zpracovávání takového množství dat toto ohlídat?

Otázka odpovědnosti a nesení následků za případné nehody je velmi aktuální a je potřeba ji vyřešit, než budou stroje uvedeny do provozu. Je nutné si uvědomit, že umělá inteligence není sama za sebe, ale jedná se o celý sociotechnický systém, jehož součástí nejsou jen softwarové programy, ale i lidé, instituce a organizace.<sup>110</sup>

---

<sup>107</sup> Srov. SIMON, Judith. „Distributed Epistemic Responsibility in a Hyperconnected Era“. In: *The Onlife Manifesto: Being Human in a Hyperconnected Era*. Edited by FLORIDI, Luciano. Imprint: Springer, 2015. 145.

<sup>108</sup> Srov. COECKELBERGH, Mark. *Etika umělé inteligence*. Filosofía, nakladatelství Filozofického ústavu AV ČR, v.v.i., 2023. 142.

<sup>109</sup> Tamtéž, 143.

<sup>110</sup> Srov. DIGNUM, Virginia. „Responsibility and artificial intelligence“ In *The Oxford Handbook of Ethics of AI*, edited by DUBBER, D. Markus, PASQUALE, Frank, DAS, Sunit. Oxford University Press, 2020. 215.

Níže navrhuji etické zásady pro bezpečnost a ochranu před zneužitím:

- 1) Do vývoje musí být začleněni lidé z co nejvíce sociálních skupin.
- 2) Při návrhu UI by se mělo dbát na to, aby autonomní systémy pracovaly i s takovými daty, jako jsou společenské hodnoty, morální a etické zásady, a rozhodování by se mělo odvíjet na základě těchto dat.
- 3) Všechny zúčastněné strany (vývojáři, výzkumníci, designéři, manažeři, pracovníci v nákupu a prodeji) by měly absolvovat školení, pochopit svůj úkol v celkovém procesu a převzít za část své práce zodpovědnost.
- 4) Stejně tak by všechny zúčastněné strany měly být seznámeny s možnými dopady a důsledky jejich práce na společnost.
- 5) UI by měla být naprogramována tak, aby bylo možné dohledat, proč došla k určitému rozhodnutí.
- 6) Veškeré vývojové části programu by měly být transparentní a dohledatelné.
- 7) UI systémy by měly být naprogramovány tak, aby činily rozhodnutí, která bychom považovali za etická, pokud bychom je učinili jako lidé (neboli UI by měla dodržovat Kantův kategorický imperativ).
- 8) Nedílnou součástí při práci s daty a vývojem UI je participace aneb nutné porozumění tomu, jak různí lidé pracují a žijí s technologiemi UI napříč kulturami<sup>111</sup>.

Tato doporučení spadají pod následující etické principy: *princip beneficence, princip non-maleficence, princip transparentnosti, princip spravedlnosti, princip odpovědnosti*.

### **3.1.5 Udržitelnost a environmentální ohleduplnost**

Jak je vidno z předešlých odstavců a problematických pasáží, umělá inteligence je především navrhována s orientací na člověka.<sup>112</sup> Jak je to tedy s orientací na přírodu, živočichy a planetu vůbec? Nyní žijeme v období antropocénu, kdy lidé svou činností ovlivňují ekosystém planety.<sup>113</sup> Jaký dopad bude mít v tomto ohledu technologie? Nejdříve je nutné se podívat na vztah člověka k životnímu prostředí (dále také „ŽP“) a dle toho následně hodnotit vztah dopadu technologií na ŽP.

---

<sup>111</sup> Srov. DIGNUM, Virginia. „Responsibility and artificial intelligence“ In *The Oxford Handbook of Ethics of AI*, edited by DUBBER, D. Markus, PASQUALE, Frank, DAS, Sunit. Oxford University Press, 2020. 215–220.

<sup>112</sup> Srov. COECKELBERGH, Mark. *Etika umělé inteligence*. Filosofía, nakladatelství Filozofického ústavu AV ČR, v.v.i., 2023. 217.

<sup>113</sup> Srov. STEFFEN, W. & CRUTZEN, PAUL & MCNEILL, JOHN. „The Anthropocene: Are Humans Now Overwhelming the Great Forces of Nature“. In: *AMBIO A Journal of the Human Environment*. 2008. v. 36, 614–621. 36.

V dnešní době už jsou programy a některé metody UI zapojeny do řešení klimatických problémů. Z té pozitivní roviny jde o výhody UI zpracovávat velká množství dat a tím dojít k řešením, která běžný člověk odhalit nemůže. Následně lze s návrhem realizovat a koordinovat opatření, například v otázce snížení škodlivých emisí.<sup>114</sup>

Dále si uvedme několik dalších příkladů dle Celine Herweijer, kde lze UI využít pro udržitelný rozvoj. Jedná se především o modelaci a monitoring na základě velkého množství dat a tím o zvýšení optimalizace ve prospěch udržitelného rozvoje. Například optimalizace u autonomních a propojených elektrických vozidel, kde díky optimalizaci tras jízdy a nastavení ekologických jízd dochází ke snížení skleníkových plynů. Dále distribuované energetické sítě, kde opět díky velkému množství dat, která pomocí práce UI stanoví poptávku a nabídku, lze efektivně řešit skladování a výrobu energií. A v neposlední řadě inteligentní zemědělství a potravinové systémy, kde opět díky práci UI s nasbíranými daty je možné odhalit případné problémy a choroby plodin, zvířat a také zefektivnit výrobu na základě dat poptávky a nabídky. To vše může vést ke zvýšení efektivity využívání zdrojů v zemědělství, snížení spotřeby vody, hnojiv a pesticidů a zvýšení odolnosti klimatu.<sup>115</sup>

Negativním dopadem systémů UI je především velká spotřeba energie a též využití materiálů, které jsou potřeba pro výrobu hardwarů, jako jsou čipy, telefony a další. Umělá inteligence pracuje s miliardou dat a výpočetních systémů, které spotřebují velké množství energie. Jedná se o dlouhé a náročné výpočty, ke kterým je potřeba mnoho technologie. Roy Schwartz uvedl termíny jako Red AI a Green AI systémy. Jedná se o systémy, které jsou využívány ve výzkumu UI. Zatímco Red AI se zaměřuje na přesnost výsledku, tak Green AI na její efektivitu. Oba systémy jsou pro vývoj důležité s tím rozdílem, že Red AI pracuje s velkým množstvím dat a velkým množstvím přepočítávání, čímž spotřebovává více energie, ale i finančních prostředků.

---

<sup>114</sup> Srov. FLORIDI, L., COWLS, J., BELTRAMETTI, M., CHATILA, R., CHAZERAND, P., DIGNUM, V., LUETGE, Ch., MADELIN, R., PAGALLO, U., ROSSI, F., SCHAFER, B., VALCKE, P., VAYENA, E. „AI4People-an Ethical Framework for a Good AI Society: Opportunities, Risks, Principles and Recommendations,“ In: *Minds and Machines*, 2018, 28, č. 4, 689–707.

<sup>115</sup> Srov. HERWEIJER, C. „8 Ways AI Can Help Save the Planet“. In: *World economic forum*. [online]. 13.8.2023]. Dostupné z: <https://www.weforum.org/agenda/2018/01/8-ways-ai-can-help-save-the-planet/>

Zatímco Green AI se zaměřuje hlavně na efektivnost a možnost snížení výpočtů s cílem dosažení výsledku. Green AI se vzhledem k nedostatečným finančním prostředkům používá zejména ve školách a státních výzkumných centrech, zatímco Red AI využívají více soukromé subjekty.<sup>116</sup> Ve vývoji UI jsou potřebné oba systémy, jež by měly spolu kooperovat.

Jakým způsobem tedy navrhovat UI s ohledem na životní prostředí? Měly by být v tomto případě zohledněny více hodnoty člověka, nebo planety? Vezměme si příklad s autonomním vozidlem, kde bude umělá inteligence navrhovat jízdu tak, aby byla co nejekologičtější. Do jaké míry toto řízení zasahuje do autonomie člověka? A co je v této situaci správné?

Níže navrhuji etické zásady pro vývoj „udržitelné umělé inteligence“:

- 1) UI by měla mít ve svých datech zakořeněné i hodnoty dotýkající se udržitelného rozvoje.
- 2) Systémy Green AI a Red AI by spolu měly kooperovat, aby vznikl systém, který bude přesný a zároveň co nejméně ekologicky náročný.
- 3) Všechny zúčastněné strany (vývojáři, vědci, designéři, manažeři, pracovníci v nákupu a prodeji) by měly být školeny i v oboru udržitelného rozvoje a dopadu UI na planetu.
- 4) Klimatické problémy se nesmí omezit jen na technologickou sféru, ale musí být dále řešeny i na poli politickém a sociálním.

Uvedené etické zásady spadají pod následující etické principy: *princip beneficence*, *princip non-maleficence*, *princip spravedlnosti*.

### **3.1.6 Inovace a společenská prospěšnost**

U umělé inteligence stejně jako u jiných technologií dochází k neustálým inovacím. Inovace v oboru UI probíhá velice rychle a často předbíhá ostatní obory, které by s tímto vývojem měly být spojeny, jako jsou právě etika, politika, ekonomie, právo a další.

Uvedme si příklad automatizace výrobních linek. Tato inovace je na jedné straně velice prospěšná pro ekonomii a rozvoj, ale na druhé straně negativně dopadá na zaměstnanost navíc stále více rozevívá nůžky mezi chudými a bohatými lidmi. Díky automatizaci lze dosáhnout vyšší produktivity, což následně snižuje cenu zboží a podporuje konzumní styl života. Zde si pak lze klást otázku: „Kdy bude mít lidstvo

---

<sup>116</sup> Srov. SCHWARTZ, R., DODGE, J., SMITH, N. A., & ETZIONI, O. „Green ai“. In: *Communications of the ACM*, 2020. 63(12). 54–63.



dost?“ Zde není možná ani potřeba tolik objektivizovat, stačí, když si čtenář subjektivně odpoví. Tento způsob pak ovlivňuje jak myšlení celé společnosti, tak dopadá na ekologickou sféru. Pozitivní stránkou je pak větší rozmanitost produktů a výrobky rozsáhlé škály kvality. Negativním dopadem je fakt, že o práci přicházejí méně kvalifikovaní pracovníci, zatímco více vzdělaní a kvalifikovaní lidé tímto více vydělávají. Tento problém je potřeba řešit hned v zárodku a komplexně.<sup>117</sup>

S inovacemi a návrhem UI pro společenskou prospěšnost je spojen pojem *Design Thinking*. Jedná se o proces, který se zabývá inovací produktů, služeb a procesů, a to s ohledem na potřeby zákazníka. Řeší tyto tři okruhy: zda je požadavek technicky proveditelný, ekonomicky životaschopný a žádoucí pro uživatele.<sup>118</sup> Design Thinking dále pracuje s těmito zásadními etapami: empatie, definování problému, generování nápadů, tvorba prototypu a testování.<sup>119</sup> Problémem tohoto postupu je například zaměření čistě na konkrétního zákazníka, čímž pak mohou vyvstávat problémy s diskriminací či neudržitelností a jednostranností. Dalším problémem je zaměření vývoje pouze na vývojáře a jeho skupinu, nikoliv na odborníky zabývající se touto problematikou. Inovace je prospěšná, pokud bude probíhat ve všech odvětvích, a to v souladu s ostatními odborníky a obory. Inovace by se neměla zaměřovat pouze na sféry, které generují zisk, ale i na sféry sociální, politické, zdravotní a ekologické.

Níže navrhuji etické zásady k inovaci a společenské prospěšnosti:

- 1) Inovace by měla počítat s rozmanitostí a různorodostí hodnot všech lidí.
- 2) Inovace UI by neměla probíhat pouze u systémů, které generují primárně zisk.
- 3) Vývojáři by měli své postupy konzultovat s odborníky z daného oboru.
- 4) Vývoj UI by měl postupovat tak, aby byl ruku v ruce s etikou, právem, politikou, ŽP aj.
- 5) Inovace technologií by neměla negativně ovlivňovat životní prostředí.

Navržené etické principy k tématu inovace a společenské prospěšnosti, které plynou ze zásad, jež jsem navrhla: *princip beneficence, princip non-maleficence, princip spravedlnosti, princip odpovědnosti*.

---

<sup>117</sup> Srov. BRYNJOLFSSON, Erik. *Druhý věk strojů: práce, pokrok a prosperita v éře špičkových technologií*. Přeložil Filip DRLÍK. V Brně: Jan Melvil Publishing, 2015. Pod povrchem. Poz. 506.

<sup>118</sup> Srov. ARNOLD, John E. *Creative Engineering: Promoting Innovation by Thinking Differently*. Edited with an introduction and biographical essay by William J. Clancey. Stanford Digital Repository. Retrieved 13 December 2016.5.

<sup>119</sup> Srov. ENGHOLM, Ida. *Quick guide to design thinking*. 1. vyd. Copenhagen.110.

## 3.2 Mravní zodpovědnost uživatele UI

V předešlé kapitole byly představeny etické aspekty pro aktéry, kteří se zabývají vývojem umělé inteligence. Jak bylo popsáno, nejedná se pouze o vývojáře, ale také o investory, kteří zadávají své požadavky na funkce UI. V nemalé míře je také potřeba věnovat se etickému chování osob, které stojí na výstupní straně programu, a to jsou uživatelé. Je potřeba vycházet z již uvedené premisy, že umělá inteligence funguje na základě strojového učení, tedy i na základě toho, jak s ní uživatel pracuje a jaké informace jí dodává či požaduje. Je možné říci, že ve fungování UI se částečně odráží chování a fungování uživatelů, tedy nás všech. Uživatele lze pak rozdělit na běžné uživatele, což je většina lidí ve společnosti, a odborné uživatele z oblastí lékařství, kriminalistiky, práva, armády, politiky, ekonomie a další.

### 3.2.1 Práce s daty ze strany uživatele

Stejně jako v předešlé kapitole je jednou z nejdůležitějších věcí při nakládání s UI práce s daty a ochrana soukromí. Zatímco v předešlé kapitole jsme se zaměřili na ochranu dat z pohledu vývojářů, zpracovatelů a majitelů dat, v této kapitole se budeme věnovat etickým zásadám pro práci s daty ze strany uživatele. Problematické otázky u práce s daty ze strany uživatele vyvstávají v zadávání dat do aplikací a systémů UI a dále v nakládání s nimi. Nejedná se pouze o běžné uživatele internetu a aplikací, ale i odborné uživatele, kteří systémy UI ve svých oborech využívají.

Navrhuji rozdělení dat vkládaných uživatelem do těchto čtyř skupin:

- 1) Data obsahující identifikaci, zdravotní záznamy či bankovní záznamy uživatele.
- 2) Veřejný sběr dat.
- 3) Dobrovolná data o aktivitách.
- 4) Dobrovolná data k marketingovým účelům.

*1) Data obsahující identifikaci, zdravotní záznamy či bankovní záznamy uživatele.*

Do této oblasti patří data, která obsahují naše identifikační údaje – jméno a příjmení, rodné číslo, datum narození, číslo OP nebo také data ze zdravotních záznamů či bankovní účty a karty. Uložení a zpracování podléhá vysoké ochraně a uživatel by měl zvažovat každé zadání těchto dat do systému. Dále by měl dbát na vysokou míru ochrany svých systémů, jelikož úměrně s vývojem technologie roste i vývoj hackerských aplikací a přibývá hackerských útoků. Uživatel by tak neměl spoléhat pouze na ochranu ze strany správce dat, ale také na svou osobní ochranu. Například pokud nemá uživatel počítač zabezpečený antivirovým programem, může dojít k napadení jeho interních informací. Uživatelův systém může také vytvořit pomyslnou

bránu do banky dat. Zde je pak ekonomickou otázkou, do jaké míry má být zajištěna antivirová ochrana ze stran zprostředkovatele a do jaké míry ze strany uživatele. Pokud bude narůstat požadavek na antivirovou ochranu na zprostředkovatele, vzniknou náklady na její vývoj a aplikaci a uživatel bude platit za každý specifický program větší částku. Popřípadě se může tato zodpovědnost přenechat na koncovém uživateli, ale zde je pak kladen velký důraz na dodržování antivirové ochrany.<sup>120</sup>

V dnešní době narůstá počet podvodů v bankovní sféře a ve sféře populárních online nákupů. Podvodník vytvoří shodnou webovou stránku obchodu či banky a přiměje zákazníka se přihlásit či zadat platební údaje, tím se pak dostane k jeho účtům. Banky a obchody už na tyto podvody reagují a upozorňují na klamavé reklamy a možné záměny. Úkolem uživatele je informovat se o těchto podvodech a nejlépe varovat další uživatele ve svém okolí, kteří se v technologiích tolik neorientují.

Další etickou otázkou je zpracování těchto dat odborným uživatelem. Umělá inteligence je hojně využívána v medicíně, kde jsou sbírány lékařské záznamy a materiály. Na jejich základě lze díky UI provádět diagnostiku z daleko širšího spektra dat, než mohou mít lékaři sami. Například UI od firmy DeepMind má ve své databázi miliony snímků z očních skenů a na základě těchto dat se naučila diagnostikovat známky degenerativních onemocnění očí. Další systémy se na základě radiologických snímků učí diagnostikovat rakovinu či infekční choroby, a dokonce stanovit postup léčby.<sup>121</sup> K tomu, aby mohla být data využívána, musí lékař požádat o informovaný souhlas pacienta. Informovaným souhlasem v tomto konkrétním případě je myšleno, že uživatel (pacient) bude obeznámen s fungováním použitého systému.<sup>122</sup> Lékař nesmí data zneužívat a šířit je mimo povolené databáze či ke své osobní potřebě. Tato data mohou na jednu stranu být velkým přínosem, ale nelze zapomínat ani na lidskou inteligenci a intuici. Systémy UI mají sloužit pouze jako pomocná síla, nikoliv jako hlavní aktér. Lékař by měl vždy ověřit dostupné zdroje a výsledky a následně prověřit výsledek umělé inteligence z hlediska své zkušenosti a intuice.

Každý, kdo ve svém oboru pracuje s těmito daty, musí respektovat etické zásady práce s nimi, jejich ochranu a soukromí uživatele.

---

<sup>120</sup> Srov. VARIAN, H. *Intermediate Microeconomics-A Modern Approach*. páté vydání, WW Norton and Company, New York, 1999. 105.

<sup>121</sup> Srov. COECKELBERGH, Mark. *Etika umělé inteligence*. Filosofie, nakladatelství Filozofického ústavu AV ČR, v.v.i.,2023. 116.

<sup>122</sup> Srov. ANOM, B.Y. „Ethics of Big Data and artificial intelligence in medicine.“ In: *Ethics, Medicine and Public Health*. [online]. 2020, roč. 15. 10. [cit. 2024-04-01].

Dostupné z: <https://doi.org/10.1016/j.jemep.2020.100568>.

### 1) *Veřejný sběr dat – bez souhlasu*

Sběr těchto dat probíhá bez souhlasu uživatele. Jedná se například o data z bezpečnostních nebo dopravních kamer. Běžný uživatel může požádat o informace, jak jsou sbíraná data zpracovávána a uchovávána. Odborný uživatel, který data zpracovává, musí dodržovat etické zásady a respektovat ochranu a soukromí lidí. Jak bylo zmíněno v předešlé kapitole, často jsou data používána například v soudních procesech nebo predikci policejních zásahů a rozmístění hlídek. Stejně jako v případě lékařů je důležité si uvědomit, že se jedná o data UI, která pracuje s velkým množstvím dat, a proto – jak bylo popsáno v předešlé části 3.2.2. *Rovnost a nediskriminace* – mohou být tato data mohou často neobjektivní. Například při soudním procesu konkrétního člověka v případě systému KOMPAS. Pomocí tohoto systému UI zpracovává závěry na základě dostupných dat a určuje, kdo je předurčen k tomu, aby trestný čin spáchal znovu a kdo ne.<sup>123</sup>

### 2) *Dobrovolně vkládaná data o aktivitách*

Dalším typem vkládaných dat je dobrovolné vkládání dat o aktivitách. Žijeme v době digitální a veškeré naše činnosti jsou spojeny i se světem internetu. Mnoho aplikací, systémů a domácích robotů sbírá data o našich fyzických aktivitách, denních rutinách, dovolených, kulturních akcích a vcelku o všem, co děláme. Jedná se především o chytré telefony, které máme při většině aktivit při sobě a používáme je k vyhledávání téměř neustále. Také bylo vyvinuto mnoho aplikací, které nám nabízejí kontrolu našeho zdraví, měření tepu, kontrolu spánku aj. Na jedné straně jsou pro nás tyto aplikace prospěšné, šetří nám čas, motivují nás k lepším výsledkům (například sportovní a učící aplikace), upozorňují nás na problémy a v některých případech mohou zachránit i život. Uveďme si příklad chytrých hodinek, které při neobvyklém pohybu – zakolísání při běhu – dají uživateli otázku, zda je vše v . a pokud ne, samy zavolají SOS. Na straně druhé je potřeba si uvědomit, že je tímto náš život zaznamenáván a ukládán v digitálním světě. Další otázkou je, do jaké míry se uživatel v digitálním světě rozhoduje autenticky a do jaké míry je jeho rozhodování ovlivňováno UI neboli tzv. nudging, kterému se budeme věnovat v další kapitole 3.2.2 *Ovlivňování, manipulace a zneužití UI*.

---

<sup>123</sup> Srov. COECKELBERGH, Mark. *Etika umělé inteligence*. Filosofie, nakladatelství Filozofického ústavu AV ČR, v.v.i., 2023. 155.

Tato data jsou v převážné většině využívána dále a prodávána třetím stranám k marketingovým účelům, jako je to například u platformy Facebook či Instagram.<sup>124</sup>

Co se týká soukromí a ochrany dat, zmiňované aplikace, systémy a roboti vždy vyžadují souhlas s užíváním dat a ochranou soukromí. Upřímně – jen malé procento uživatelů tyto rozsáhlé smlouvy čte, jelikož využívaných aplikací je nespočet a tím pádem i nespočet smluv. Zpracování dat je však transparentní, a záleží tedy pouze na rozhodnutí uživatele, zda dané smlouvy čte. Uživatel si také často neuvědomuje, že souhlas dává pouze na začátku používání aplikace, např. při koupi chytrého telefonu od firmy Apple a spolupráci s virtuálním asistentem SIRI. Ve smluvních podmínkách podepisujeme ukládání a zpracování dat při používání asistenta, nikoliv při každém úkonu zvlášť. SIRI už nevyžaduje souhlas ke každému konkrétnímu kroku. Proto je potřeba být prozíraví a zvažovat, jaká data s digitálním světem sdílíme. Šedou zónou této problematiky je pak mladší a starší generace, která se nezamýšlí nad možnými důsledky a vlivem systémů na jejich osobnost, ale zaměřují se na nabízené výhody produktu. U mladší generace se jedná například o dětské hračky jako panenka Hello Barbie, která dokáže komunikovat s dětmi a ukládá data těchto rozhovorů na serverech společnosti ToyTalk.<sup>125</sup>

### 3) „Dobrovolně“ vkládaná data k marketingovým účelům

V této části se jedná především o data, která uživatel zadává při výběru zboží a služeb. Uživatel vyplňuje dotazníky o svých preferencích, vyhledává a nakupuje zboží přes internet. Veškerá data jsou pak zaznamenávána a dle zadaných parametrů a vyhledávaného zboží je pak nabízen podobný produkt. Vkládáním těchto dat se vytváří o uživateli digitální obraz, a to nejen o kupovaných produktech, ale data jsou následně zpracovávána ve sféře ekonomické, politické i sociální.

Níže navrhuji etické zásady pro práci s daty:

- 1) Uživatelé by se měli vždy informovat o fungování dané aplikace, systému či robota a o zpracovávání a sběru dat.
- 2) Uživatelé by měli informovat své okolí ohledně možného zneužití dat především ty, kdo spadají do ohrožených skupin.

---

<sup>124</sup> Srov. MARR, Bernard. „27 Incredible Examples of AI and Machine Learning in Practice“, In: *Forbes*. [online]. 30. 4. 2018. 10. [cit. 2023-04-01].

Dostupné z: <https://www.forbes.com/sites/bernardmarr/2018/04/30/27-incredible-examples-of-ai-and-machine-learning-in-practice/#6b37edf27502>

<sup>125</sup> Srov. COECKELBERGH, Mark. *Etika umělé inteligence*. Filosofía, nakladatelství Filozofického ústavu AV ČR, v.v.i., 2023. 116.

- 3) Uživatelé by si měli vždy ověřit zdrojové stránky, než zadají své identifikační údaje či jiná privátní data.
- 4) Uživatelé by se měli zaměřit na antivirovou ochranu svých programů.
- 5) Odborní uživatelé, kteří mají přístup k databázím, nesmí data nijak zneužívat.
- 6) Odborní uživatelé by měli výsledky UI vždy zkontrolovat a posoudit dle svých zkušeností a znalostí.
- 7) Odborní uživatelé, kteří vyžadují informovaný souhlas, musejí poskytnout přesné a ucelené informace.

Mnou výše navržené etické zásady pro práci s daty spadají pod tyto etické principy: *princip beneficence, princip non-maleficence, princip autonomie, princip spravedlnosti, princip transparentnosti, princip odpovědnosti.*

### **3.2.2 Ovlivňování, manipulace a zneužití systémů UI**

Velkým nebezpečím systémů umělé inteligence je jejich zneužití a s tím související ovlivňování a manipulace. Tato hrozba je velice rozsáhlé téma a určitě by vydala na samostatnou práci. V této části si uvedeme pouze některé příklady a krátká vysvětlení. Následně si představíme zásady, jak je možné pracovat s umělou inteligencí, abychom se jako uživatelé tohoto vyvarovali.

Technologové, podnikatelé a politici dnes využívají vědecké informace z kognitivní vědy, psychologie a behaviorální ekonomie k tomu, aby vytvořili programy, jejichž prostřednictvím lze s uživateli manipulovat.<sup>126</sup> Někdy si ani uživatel přímo nevšimne, kolik věcí je mu našeptáváno a náhodně představováno. Tomuto jevu, kdy je uživateli drobnými změnami ovlivněno jeho rozhodnutí, se říká „nudging.“ Tento jev vychází od anglického slovesa nudge, tedy v překladu pošťouchnout.

Nejprve si uvedme, jak takové systémy manipulace fungují a jak se používají. Představme si dva vybrané psychologické mechanismy, které se využívají v marketingu a nudgingu. Za prvé se jedná o efekt dotace, což znamená, že lidé se hůře loučí s věcmi, které už mají, než když je v ruce vůbec neměli.<sup>127</sup> Ve své knize *Pojednání* David Hume píše: „Lidé obvykle upínají své city spíše k tomu, co vlastní, než k tomu, co nikdy

---

<sup>126</sup>Srov. FRISCHMANN, B., & SELINGER, E. *Re-Engineering Humanity*. Cambridge: Cambridge University Press, 2018. 6.

<sup>127</sup>Srov. BOVENS, Luc. „The Ethics of Nudge“. In: Mats J. Hansson & Till Grüne-Yanoff (eds.), *Preference Change: Approaches from Philosophy, Economics and Psychology*. Berlin: Springer, Theory and Decision Library A. Chapter 10. 2008. 207.

neměli...“<sup>128</sup> A za druhé – lidé jsou pevnější ve svých rozhodnutích, pokud se jedná o věci budoucí než o současné náklady a užitky.

Luc Bovens pak ve svém pojednání *The ethics of Nudge* uvádí šest typů dějů, které mohou být příkladem nudge:

1) *Nevědomost*

Nevědomost je jev, kdy máme ve svém jednání jasný cíl, ale prostředky k dosažení cíle vyžadují odborné znalosti, které nám scházejí. Tudiž jsme odkázáni na doporučení odborníků. Jedním z příkladů je doporučená léčba u lékaře nebo stanovení plánů pro penzijní spoření.<sup>129</sup> V případě užívání umělé inteligence se s tímto jevem setkáváme také. Například z neznalosti práce s daty konkrétní společnosti pouze využíváme vyplývající pozitiva, ale už nevidíme nástrahu ve formě úpravy algoritmu tak, aby byl zákazník nabádán ke konkrétnímu výběru či řešení.

2) *Setrvačnost*

Jedná se o stav, kdy využíváme nabízené systémy kvůli nedostatku času a setrváváme v systému, který stále více rozhoduje za nás.<sup>130</sup>

3) *Akrasia – nevoli – slabá vůle*

Při dosahování dlouhodobých cílů máme slabou vůli, proto často podléháme nátlaku okolí a nabízené reklamě.<sup>131</sup>

4) *Podrážděnost*

Luc Bovens popisuje podrážděnost jako stav, kdy se danými věcmi nechceme zabývat.<sup>132</sup> V užití UI je to možné uvést na příkladu, kdy si řekneme, že se nebudeme zabývat tím, co se s našimi daty děje, a budeme rovinnou soukromí a autenticity ignorovat.

Poslední dva typy dějů *výjimka* a *sociální dávky* se nevztahují k tématu umělé inteligence, ale jedná se spíše o sociální stránku.

Nyní si níže uvedme některé konkrétní příklady ovlivňování, manipulace a zneužití při použití umělé inteligence. Je potřeba si uvědomit, že výše zmíněné jevy souvisí jak s naším využíváním umělé inteligence, tak se zneužíváním údajů ostatních lidí. Pokud si

---

<sup>128</sup> HUME, D. *A Treatise of Human Nature*. (2. vydání). Selby-Bigge, Los Angeles, Oxford: Clarendon, 1978.

<sup>129</sup> Srov. BOVENS, Luc. „The Ethics of Nudge“. In: Mats J. Hansson & Till Grüne-Yanoff (eds.), *Preference Change: Approaches from Philosophy, Economics and Psychology*. Berlin: Springer, Theory and Decision Library A. Chapter 10. 2008. 210.

<sup>130</sup> Srov. BOVENS, Luc. „The Ethics of Nudge“. In: Mats J. Hansson & Till Grüne-Yanoff (eds.), *Preference Change: Approaches from Philosophy, Economics and Psychology*. Berlin: Springer, Theory and Decision Library A. Chapter 10. 2008. 210.

<sup>131</sup> Tamtéž. 210.

<sup>132</sup> Tamtéž. 210.

stanovíme své hranice, které méně respektují naše soukromí a autenticitu rozhodování, není možné toto rozhodnutí dělat za druhé. Jedním z příkladů je kontroverzní situace na univerzitě Oral Roberts, kde měli studenti používat na hodiny tělesné výchovy fitness náramky „Fitbit,“ které počítaly studentovi počet kroků a fyzickou aktivitu během celého roku. Tato sledovaná aktivita představovala až 20 % ze známky studenta.<sup>133</sup> I přestože ohledně užívání náramků podepisovali studenti informovaný souhlas, dochází zde k určitému druhu donucování. Samozřejmě bylo na volbě studenta, zda tento typ využije, ale vzhledem k tomu, že to do velké části ovlivňovalo celkovou známku, o jistý nátlak bezesporu šlo. Dalším příkladem může být použití chytrého systému či aplikace lidmi v důchodovém věku. Opět může být záměr zcela pozitivní a ochranný, je však potřeba s tím dotyčného seznámit. Stejně tak bychom měli dbát na to, abychom bez souhlasu jiného člověka nezveřejňovali jeho údaje, přestože se může zdát banální označit kamaráda na výletě na sociálních sítích. Tím však už rozhodujeme o šíření jeho aktivity v digitálním světě.

#### *Online nákupy*

Jedním z odvětví, kde se hojně využívají manipulativní techniky za pomoci sběru dat a algoritmů UI, je online nákup zboží. Určitě to každý zná, že pokud vyhledává konkrétní typ zboží, minimálně další týden jsou mu nabízeny reklamy na tento typ zboží. Tato podprahová reklama je velmi manipulativní, nabízí nám zboží, které algoritmus umělé inteligence na základě námi vkládaných dat vyhodnotí jako pro nás správný. Dále nás reklama podněcuje k nákupu podobných produktů. Zde je pak na uživateli, do jaké míry se těmito systémy řídí a do jaké míry si zachová svůj vlastní úsudek. Problémem opět je, že tyto systémy většinou míří primárně na skupiny lidí, kteří mají vůli slabší, popřípadě se jedná o děti nebo starší osoby.

Dalším palčivým problémem je diskriminace ostatních dodavatelů na trhu. Vezměme si příklad při nakupování potravin online, nebo nakupování přes „chytrou“ lednici, která rovnou vypíše uživateli seznam nákupu. Uživatel zadá seznam do online obchodu a lednici zaplní. Problém nastává, když UI vytváří na základě těchto dat svůj okruh dodavatelů a tím diskriminuje ostatní. Stejně jako u předchozích aplikací má i toto dvě strany mince. Uživateli to na jedné straně často pomáhá, jelikož se mu zrychlí výběr a nabídne rovnou požadovanou věc. Na druhé straně brání uživateli, aby našel jinou

---

<sup>133</sup> Srov. FRISCHMANN, B., & SELINGER, E. *Re-Engineering Humanity*. Cambridge: Cambridge University Press, 2018. 21.



alternativu či měl svobodu v rozhodování. Jako uživatelé bychom tedy měli mít na paměti, že je v pořádku tyto systémy využívat, ale vždy zapojovat i své myšlení a být si vědomi toho, že jsou i jiná alternativní řešení a že to, co nabízí umělá inteligence, nemusí být pro nás vždy nejlepší.

### *Politická manipulace*

Následujícím typem, který si uvedeme, je politická manipulace. Před prezidentskými volbami v USA v roce 2020 bylo dokázáno, že průběh voleb ovlivňovala data ze sociální platformy Facebook, která sbírala data od uživatelů vyjadřujících se k politické situaci.<sup>134</sup> Stejně tak je ve vzduchu hrozba před budoucími volbami v roce 2024. UI na základě dat například vytvoří statistiku preferencí a tyto statistiky následně zveřejní.

### *DeepFake*

Dalším z obávaných manipulativních nástrojů je tzv. „DeepFake“. Jedná se o vytváření klamných videí, audiozáznamů a obrazů za pomoci umělé inteligence, která pracuje na základě hlubokého učení a rozsáhlé neuronové sítě.<sup>135</sup> Vytváření obrazových materiálů je nyní dostupné od umělé inteligence např. OpenAI, tzv. DALL.E, popřípadě DEEPLY.CZ. Za menší poplatek lze vytvořit na základě zadaných slov v podstatě reálný obraz. Umělá inteligence je tímto pak zneužívána pro vytváření tzv. FakeNews, tedy dezinformací. Základem práce s informacemi na internetu je tedy ověřování dat a informací.

A v neposlední řadě je potřeba si uvědomit, že prací s UI ovlivňujeme její výsledky všichni. Jelikož se umělá inteligence učí nejen z trénovací databáze dat, ale následně i z dat zadaných uživateli a pracuje následně s daty nás všech a dle toho pak vytváří odpovědi. Opět si uveďme příklad chatbota, který byl v roce 2016 od firmy Microsoft uveden na trh. Chatbot Tay byl spuštěn na Twitteru a následně musel být odstraněn, jelikož mu byla uživateli Twitteru zadávána rasistická data a chatbot poté začal přidávat též rasistické komentáře.<sup>136</sup>

Na základě výše uvedeného je pro uživatele důležité si uvědomovat, jaká data do systému zadávají a jaká vyhledávají. Jednou rovinou je tedy etika uživatelů, kteří jednají vědomě a zneužívají umělou inteligenci ke svému prospěchu, a druhou je rovina

---

<sup>134</sup> Srov. COECKELBERGH, Mark. *Etika umělé inteligence*. Filosofía, nakladatelství Filozofického ústavu AV ČR, v.v.i., 2023. 128.

<sup>135</sup> GPT-3.5, „Deepfake je složenina slov ‚deep learning‘ (hluboké učení) a ‚fake‘ (falešný)...“, [4. 9. 2023].

<sup>136</sup> Srov. COECKELBERGH, Mark. *Etika umělé inteligence*. Filosofía, nakladatelství Filozofického ústavu AV ČR, v.v.i., 2023. 129.

všech uživatelů, kteří zadávají data a tím spolupracují na vývoji daného algoritmu. Systémy UI budou chtít nechtít součástí naší budoucnosti a tím, že zadáváme svá data, můžeme mít nepřímo vliv na ostatní uživatele. Do systémů UI se odráží stav naší společnosti, a pokud tedy z ní máme strach, zaměříme se v první řadě na své chování.

Brett M. Frischmann a Evan Selinger v knize *Re-Engineering Humanity* uvádějí: „*Boj za svobodné bytí v offline světě bude jedna z nejvíce důležitých bitev jednadvacátého století.*“<sup>137</sup> Etické zásady by měly být směřovány na aktéry, kteří manipulativní obsahy za pomoci UI vytvářejí, jako jsou již zmiňovaní technologové, podnikatelé, politici a další uživatelé. A na druhé straně by měli sami uživatelé umět s obsahem informací pracovat tak, aby si manipulace byli vědomi a snažili se udržovat si svá vlastní rozhodnutí.

Níže navrhuji etické zásady při práci s UI v souvislosti se zneužíváním, manipulací a ovlivňováním:

- 1) Uživatel by měl dbát na soukromí a ochranu dat ostatních osob.
- 2) Uživatel by neměl používat donucovací prostředky k používání chytrých nástrojů (viz náramek Fitbit).
- 3) Uživatel by měl být na pozoru ohledně vlivu manipulace a dbát na to, aby manipulaci nešířil dále.
- 4) Uživatel nesmí využívat umělou inteligenci k dezinformačnímu materiálu a šířit ho dále.
- 5) Uživatel by si měl ověřovat informace z více zdrojů, aby omezil dopady FakeNews.
- 6) Uživatel by si měl být při užívání UI vědom své odpovědnosti za data, která do UI zadává.

Mnou výše navržené zásady spadají pod následující etické principy: *princip beneficence, princip non-maleficence, princip spravedlnosti, princip transparentnosti, princip odpovědnosti, princip autonomie.*

---

<sup>137</sup> FRISCHMANN, B., & SELINGER, E. *Re-Engineering Humanity*. Cambridge: Cambridge University Press, 2018. 7.

### 3.2.3 Závislost a zodpovědnost

V textech o umělé inteligenci jsme se dostali k etickým otázkám, jako jsou zodpovědnost, diskriminace, ochrana dat, bezpečnost, ale nelze zapomenout ani na ochranu našich hodnot. Není možné si nevšimnout, jak se zvyšuje nárůst psychických onemocnění, zejména úzkostí, depresí a syndromů vyhoření. To vše úzce souvisí s tím, že se člověk odpojuje od svého bytí a svých hodnot. Proto je potřeba, aby uživatelé vždy zvážili, do jaké míry je pro ně „soužití“ se systémy UI pozitivní.

V knize *Re-Engineering Humanity* je popsána problematika outsourcingu, tedy delegování fyzické, kognitivní, emocionální a etické práce na třetí stranu.<sup>138</sup> Tímto způsobem v dnešní době pracuje mnoho lidí. Jde především o zrychlenou dobu a velké množství dat, se kterým lidé pracují pomocí třetích stran. Data si nechávají vyhledávat a zpracovávat. Outsourcing uživatelům tedy ulehčuje práci a také otevírá více možností. Otázkou je, zda je tato cesta správná, popřípadě do jaké míry. Tento styl práce přináší uživateli mnoho výhod, ale samozřejmě i nevýhod, na které by měl brát zřetel. V první řadě se jedná o *pasivitu*.<sup>139</sup> Zde je otázkou, do jaké míry necháme za sebe pracovat stroje. Dostáváme se do role spíše pasivního pozorovatele než aktivního účastníka řešení problému. S tím souvisí druhý problém, a to *snížení aktivity*. Tím, že na problému nepracujeme pouze sami, snižuje se naše kontrola nad zadaným problémem a tím se snižuje pocit *odpovědnosti*. Vezměme si příklad práce s chatbotem chat GPT. Pokud zadáme otázku či požadavek do Googlu, otevře se stránka, která je zároveň zdrojem informací. Ale chat GPT takto nefunguje. Ten vygeneruje odpověď napřímo, bez jakékoliv reference či kontroly. Někdy jsou informace i nepravdivé. Uživatel poté přenáší odpovědnost na chatbota, nicméně skutečná odpovědnost je primárně na něm, jelikož jeho povinností je si zdroje ověřit. Otázka odpovědnosti s sebou pak přináší i dopad na naši osobnost. Jelikož budeme problém řešit primárně na základě třetí strany, může to mít vliv na naši psychickou stránku. Z řešení problému nebudeme mít ve finále radost a pozitivní výsledek nám nepřinese žádné kladné emoce.<sup>140</sup>

Čtvrtým problémem je *zvýšená nevědomost*.<sup>141</sup> Zde se jedná především o pochopení správného fungování systému. Často řetězcem delegování problému dojdeme k výsledku, k němuž bychom samostatně nedospěli, tudíž s ním ve výsledku

---

<sup>138</sup> Srov. FRISCHMANN, B., & SELINGER, E. *Re-Engineering Humanity*. Cambridge: Cambridge University Press, 2018. 32.

<sup>139</sup> Tamtéž, 32.

<sup>140</sup> Tamtéž, 33.

<sup>141</sup> Tamtéž, 34.

nejsme spokojeni. V řetězci delegování však zapomínáme na osobní přínos a necháváme veškerý proces na algoritmu. Zde si uveďme příklad ze strany překladatelství. Například literární překlady jsou svým způsobem samostatným uměleckým dílem. UI bude dříve či později schopna přeložit sonet také, ale s jakým dopadem? Umělá inteligence jedná pouze na základě dostupných dat, nikoliv na základě pocitů a lidských vjemů. Tímto pak můžeme ohrozit kulturu a umění jako takové.

Pátým problémem je *odstup*. Čím více práce necháme třetí straně, tedy v tomto případě umělé inteligenci, tím větší máme odstup od daného problému.

Posledním problémem je *snížená nezávislost*. V případě, že se systémy UI stanou součástí naší práce či života, může dojít k závislosti na nich, popřípadě i ke ztrátě motivace řešit problémy či zvyšovat své znalosti a dovednosti.<sup>142</sup> Uveďme si příklad na systému automatizované navigace od Maxe Pfeiffera, tedy navigace, která je aktualizována dle aktuálních dat. Příkladem je aplikace WAZE pro automobily. Tato navigace je navržena tak, aby si každý uživatel přišel na své. Například turistům je nabídnutoa, kde se nachází více památek. Důchodcům bude nabídnuta co nejkratší a nejméně rušná cesta domů. Mladým lidem pak cesta živější s větší nabídkou aktivit. Tato aplikace však skýtá mnoho dilemat, která si sám autor uvědomuje. Dilemata kopírují výše uvedené problémy. Dle Nicholase Carra tyto systémy vedou k automatizaci a tím často k ignorování kritického myšlení či zdravého rozumu. Pokud člověk pouze poslouchá aplikaci a nepřemýšlí nad cestou, může dojít i k nehodě.<sup>143</sup> Frischmann a Selinger ve své knize *Re-Engineering Humanity* uvádějí tvrzení filozofů Huberta Dreyfuse a Seana Kellyho, kteří tvrdí, že podlehnout orientaci dle GPS je „odlidšťující“, protože trivializuje umění navigace a zanechává nás bez bohatého pocitu, kde jsme a kam jdeme.<sup>144</sup> Dalším příkladem je závislost na aplikacích v rámci fyzické aktivity. Dnes uživatelé po každé fyzické aktivitě kontrolují, zda ji chytré hodinky či telefon zaznamenaly. Radost z aktivity samé je závislá na porovnání aktivity s jiným uživatelem či porovnání svých výsledků. Poté jsou uživatelé často nuceni k fyzické aktivitě danou aplikací. Zde se pak nabízí otázka, do jaké míry slouží automatizace člověku a do jaké míry se člověk stává programovatelným.

---

<sup>142</sup> Srov. FRISCHMANN, B., & SELINGER, E. *Re-Engineering Humanity*. Cambridge: Cambridge University Press, 2018. 35.

<sup>143</sup> Tamtéž, 35.

<sup>144</sup> Tamtéž, 35.

Používání všech těchto aplikací je na jednu stranu přínosné a usnadňuje to život v dnešní zrychlené době. Na druhou stranu se tím člověk odpoutává sám od sebe a je závislý na všech aplikacích, robotech a přístrojích, které mu život organizují. Pokud se pak vrátíme na začátek, není pak zásadnější otázka – „Zda se lidé nebudou chovat více jako roboti, než zda je robot schopný se chovat jako člověk?“ Popřípadě nenaplní se Kubrickovo temné proroctví? – „*Když se začneme spoléhat na počítače, aby nám zprostředkovaly pochopení světa, pak naše vlastní inteligence splyne s umělou.*“<sup>145</sup>

Stejně jako v předešlé kapitole docházím k závěru, že každý individuální jedinec by měl dbát na své vlastní hodnoty, své kritické myšlení a svůj rozum. Je potřeba, abychom dbali na rozvoj společnosti, udržitelnosti, vztahů na základě svého vlastního „bytí“, nikoliv na základě výsledků práce umělé inteligence.

Níže navrhuji etické zásady při práci s UI v souvislosti závislosti a odpovědnosti:

- 1) Uživatel by si měl vždy ověřit zdroje informací, které mu poskytne UI.
- 2) Uživatel by měl vždy postupovat tak, aby hlavní myšlenka řešení problému byla jeho vlastní, nikoliv prostředkem UI.
- 3) Uživatel by měl dbát na své potřeby a hodnoty, nikoliv na příkazy a manipulaci ze strany systémů UI.

Výše mnou navržené etické zásady se řadí pod tyto principy: *princip beneficence, princip non-maleficence, princip spravedlnosti, princip transparentnosti, princip odpovědnosti, princip autonomie.*

### **3.2.4 Etická vzdělanost a uvědomělost**

Umělá inteligence je jedním z odvětví digitálního světa. Proto je i při práci s UI nutná digitální gramotnost. Výše jsme si uvedli několik hrozeb a problémů při návrhu a užívání UI, které vycházejí zejména z neznalosti fungování těchto systémů, digitální gramotnosti obecně a etických norem.

Vzdělávání v digitálním světě by mělo být součástí školního systému, ale zároveň by nemělo docházet k donucování využívat tyto systémy jako například chytré hodinky nebo algoritmy, jež sbírají konkrétní data o užívatelích. Aktuálně jsou školní osnovy stále nastavené spíše na nedigitální dobu a více či méně se snaží přizpůsobovat. Na jedné straně by měla být ve školství zahrnuta digitální gramotnost jako taková. V aktuální době toto vzdělávání nahrazují soukromé firmy, které pak pořádají workshopy pro žáky. A na druhé straně by měl být přizpůsoben i systém výuky

---

<sup>145</sup> In: CARR, Nicholas G. *Nebezpečná měřičina: jak internet mění náš mozek: analýza stavu lidské psychiky v době digitální.* Přeložil Jaroslava PREROVSKÁ. V Praze: Dauphin, 2017. 309.

současnému životnímu stylu. Tím je myšleno, že stávající osnovy pocházejí z doby, kdy digitální svět nehrál v životě jedince takovou roli. Informace byly získávány především ve škole, popřípadě v literatuře. Tento systém byl tedy nastaven tak, že se žáci naučili co nejvíce informací a poté s nimi pracovali a rozvíjeli je ve svých dalších působištích. Zatímco dnešní doba je v podstatě úplně opačná. Informací je všude přemíra, nehledě na to, že nemusí být pravdivé, či dokonce mohou být zavádějící. A úkolem žáka by mělo být ve větší míře informace najít, zhodnotit a zpracovat a až na základě tohoto procesu se vyhodnocené informace naučit.

Vzdělávání v digitální gramotnosti se netýká samozřejmě jen škol, ale všech institucí a uživatelů systémů UI. Vzdělávání v této oblasti nebudeme v této kapitole ani diplomové práci rozvíjet, pouze si představíme odvětví, která by měla do daného vzdělávání spadat pro bezpečné a etické užívání UI.

Oblasti vzdělávání v digitální gramotnosti:<sup>146</sup>

- 1) *Informační a datová gramotnost* – Do této části spadají informace o správě a zpracování dat.
- 2) *Komunikace a kolaborace* – Tato část by měla rozvíjet komunikaci prostřednictvím digitální technologie. Zvolení správných systémů pro konkrétní účel či úkol. Dále správnost sdílení dat a zachování soukromí ostatních uživatelů. Respektovat rozmanitost uživatelů digitálního světa, ať už ze strany genderové, rasové, generační nebo kulturní.
- 3) *Tvorba digitálního obsahu* – Informace ohledně vytváření digitálního obsahu, autorských práv a licencí. Zvláště při využití systémů UI, jako je například chatbot GPT 3.5/4, je nutné dbát na zdrojování používaných informací.
- 4) *Bezpečnost* – Zde by se měli uživatelé informovat o ochraně dat a zařízení, ochraně soukromých údajů. Také si je potřeba uvědomit dopad využívání systémů UI na zdraví a psychickou stránku člověka a věnovat tomu velkou část ve vzdělávání zaměřeném na používání těchto systémů.

---

<sup>146</sup> Srov. FERRARI, A., BREČKO, B. N., PUNIE, Y. „DIGCOMP: A Framework for Developing and Understanding Digital Competence in Europe“. 2014.

Dostupné z:

[https://www.researchgate.net/publication/282860020\\_DIGCOMP\\_a\\_Framework\\_for\\_Developing\\_and\\_Understanding\\_Digital\\_Competence\\_in\\_Europe/link/561fbcf208aed8dd19402c64/download?tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19](https://www.researchgate.net/publication/282860020_DIGCOMP_a_Framework_for_Developing_and_Understanding_Digital_Competence_in_Europe/link/561fbcf208aed8dd19402c64/download?tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19)

- 5) *Řešení problémů* – Vzdělávání v konkrétních systémech UI. Při používání těchto algoritmů je možné zefektivnit jejich výkonnost tím, že bude uživatel informován o jeho fungování a používání. Například v systému chatbot GPT 3.5 je možné využívat tzv. prompty, které dokáží blíže klasifikovat požadovaný výsledek.
- 6) *Technologické kompetence* – Do této kapitoly spadá znalost hardwaru a softwaru počítače.

Níže navrhuji etické zásady při práci s UI v souvislosti s etickým vzděláváním:

- 1) Vzdělávací instituce by měly při využívání principů UI informovat o možných negativních důsledcích a upozorňovat na její nedostatky.
- 2) Vzdělávací instituce by měly podněcovat využívání UI jako jedno z řešení, nikoliv jako hlavní řešení.
- 3) Vzdělávací instituce by měly dbát na mezioborové vzdělávání a co největší komplexnost.

Výše mnou navržené etické zásady se řadí pod tyto principy: *princip beneficence*, *princip transparentnosti*.

Vzdělávání v digitální oblasti je jedním z nejdůležitějších odvětví, jímž se zabývá mnoho firem, a to od používání systémů UI po možné hrozby. Zřetel by měl být brán na to, aby nedocházelo k zaměření vzdělávání pouze na užívání konkrétních systémů, ale aby nebyla neopomíjena etická stránka užívání. Zkrátka aby práce s těmito systémy neměla negativní dopady na člověka, jeho osobnost a hodnoty. Odtud plyne závěr, že vzdělávání by mělo být mezioborové a co nejvíce komplexní.

## Závěr

V závěru práce shrnu předmětné téma v návaznosti na vyhodnocení cílů, které byly v úvodu stanoveny. Práce byla rozdělena na dvě části. Cílem první části bylo dokázat, že umělá inteligence jako taková nemůže myslet, nemá vědomí a nemůže být sama o sobě za nic zodpovědná. Tyto výsledky tedy ukázaly, že spojení „etika umělé inteligence“ není přesné, protože na umělou inteligenci jako takovou nelze vztahovat z výše uvedených důvodů etické chování. Tím se tedy naskytla otázka, s kým lze spojovat etiku v problematice umělé inteligence. Etické zásady v tomto případě lze stanovit na straně vývojářů či zadavatelů a na straně koncových uživatelů těchto systémů.

V rámci první části práce byly položeny tři otázky. První otázka zněla: „Může UI myslet jako člověk?“ Tuto otázku jsme rozebrali z více směrů analytických filozofů a nahlédli i do úvah, které myšlení strojů nevyvrací. Základním kamenem této problematiky je Descartův dualismus, který oddělil duchovní část od fyzické. A následný vývoj vědeckých disciplín poté podporoval především stránku fyzickou nebo jinou stránku, kterou lze dokázat. Mnoho filozofů zaměřených čistě na filozofii faktů připouštělo a připouští, že myšlení UI možné je. To poté dokazují možnou simulací veškerých biologických procesů člověka na strojích. Na druhé straně pak stojí John Searle, který se opírá o čtyři mentální vlastnosti a tvrdí, že není možné popřít stranu mysli a duchovna a zaměřit se pouze na stránku materialistickou. Své tvrzení dokazuje čtyřmi mentálními vlastnostmi, které souvisejí s duchovní podstatou: intencionalita, vědomí, subjektivita mentálních stavů a mentální kauzace. Dalším průlomovým tvrzením je fakt, že mysl člověka na rozdíl od fungování stroje má syntaxi, sémantiku, zatímco stroj pracuje pouze se syntaxí. Stroj může nasimulovat chování a různá jednání člověka, ale chybí mu významovost takového chování. Searle toto ukazuje na experimentu čínského pokoje, který je popsán v příslušné kapitole.

Druhá položená otázka zněla: „Může nést UI zodpovědnost za svá rozhodnutí?“ Pro nalezení odpovědi na tento dotaz jsme aplikovali zásady odpovědnosti od Aristotela z *Etiky Nikomachovy*. Aristoteles toto téma opírá o dobrovolnost jednání. Pokud subjekt jedná dobrovolně, nese za své jednání odpovědnost. Dobrovolné jednání je takové, které má hybnou příčinu uvnitř a zároveň je vědomé. V případě UI se pak dostáváme do konfliktu hned v první části, jelikož umělá inteligence je naprogramována jinou osobou a zároveň využívá data, která jsou do databáze uložena vývojářem či uživatelem, tedy



její hybná příčina je vně, tudíž její jednání není dobrovolné. Hybná síla je vnější. Proto je možné potvrdit, že umělá inteligence nemůže být zodpovědná za své činy.

Poslední položená otázka se ptala: „Má UI vědomí?“ V této kapitole jsme si představili dva přístupy ke studiu vědomí. Jedním z nich je přirozená subjektivní zkušenost a druhým dokazování vědomí pomocí vědeckých metod. Pokud se zaměříme na druhou metodu, lze připustit, že umělou inteligenci je možné naprogramovat tak, aby simulovala vědomí člověka. Ale z hlediska přirozené subjektivní zkušenosti už o vědomí u UI mluvit nelze. Umělá inteligence nemá žádnou přirozenou subjektivní zkušenost. Vždy vychází z několika zkušeností uživatelů, popřípadě vstupních dat. Zde jsme si tedy potvrdili, že umělá inteligence vědomí nemá.

Ve druhé části diplomové práce tedy navazujeme na etické zásady a ty činitele, kterých se týká v první řadě – vývojáře a uživatele. V kapitole „Člověk jako vývojář UI“ jsme se zabývali tématy jako respektování soukromí a ochrana dat, rovnost a nediskriminace, bezpečnost a ochrana před zneužitím, zodpovědnost a důsledky, udržitelnost a environmentální ohleduplnost a inovace a společenská prospěšnost. V každé kapitole byly popsány konkrétní případy a obecné pojetí této problematiky. Na konci každé kapitoly jsem navrhla etické zásady a etické principy, jimiž by se vývojáři v konkrétních případech měli řídit. Stejně tomu bylo ve druhé části „Člověk jako koncový uživatel UI“, kde jsme si představili témata jako práce s daty ze strany uživatele, ovlivňování, manipulace a zneužití systémů UI, závislost a zodpovědnost a etická vzdělanost a uvědomělost. I v této části byly ke každé kapitole rozebrány etické zásady a principy pro uživatele.

Obě části práce jsou úzce propojeny a vztahují se na ně podobné zásady a etické principy. V závěru si tedy uvedme principy, které jsou spojeny s prací s umělou inteligencí. Jedním z hlavních principů je *princip non-maleficence*, tedy neškodit. Umělá inteligence by měla sloužit především jako nástroj a neměla by být využívána k manipulacím jakékoliv formy. Dalším důležitým principem je *princip autonomie*. UI by měla být navržena tak, aby respektovala autonomii člověka. Stejně tak při práci s ní je potřeba si zachovat svou vlastní autonomii a nenechat se ovládat informacemi stroje. Při práci s umělou inteligencí by měl být brán velký zřetel na *princip spravedlnosti a rovnosti*. Umělá inteligence je často naprogramována tak, aby se zaměřila na propagaci produktu či myšlenky a omezuje tak rozhodování uživatele na dané nabízené parametry. Stejně tak by měl být systém UI naprogramován tak, aby bral zřetel na nediskriminaci a rovnost mezi různorodými sociálními a kulturními skupinami lidí.

V případě používání UI je také velice podstatné vzdělání ohledně jejího fungování a práce s ní. S tím souvisí *princip transparentnosti*. Uživatelé by měli mít vždy možnost vědět, kde jsou zadávaná data uchovávána a jak je s nimi nakládáno. Stejně tak by měly být zachovávány veškeré algoritmy systémů. S tím souvisí další princip, a to *princip odpovědnosti*. Všichni aktéři, kteří s umělou inteligencí pracují, by si měli být vědomi odpovědnosti. Vývojáři by měli odpovědně navrhovat algoritmy a uživatelé by měli být zodpovědni za to, jak pracují s daty v těchto systémech. V neposlední řadě by měla být umělá inteligence nástroj, který nejen „neškodí,“ ale také prospívá – *princip beneficence*. Je potřeba zdůraznit, že se nejedná o prospěch konkrétní skupiny lidí, ale především o prospěch všech lidí, životního a celkového prostředí kolem nás.

Dle mého názoru je v dnešní společnosti kladen velký etický důraz na vývojáře a firmy, které tyto systémy vyvíjí jako celek. Z mého pohledu by ale tato etika měla vycházet z každého konkrétního člověka, tedy jak od vývojářů, tak uživatelů. Umělá inteligence funguje na základě zadávání dat a je jen na nás, jaká data budeme do systémů zadávat a dle jakého obrazu společnosti se tyto systémy budou vyvíjet. Etika spojená s umělou inteligencí by se tedy měla rozvíjet zejména odspodu a od každého konkrétního člověka, který s ní bude, jakkoliv v kontaktu, a od jeho osobního etického nastavení.

Stejně tak si myslím, že vedle podstatných otázek ohledně strojů a jejich myšlení, odpovědnosti a vědomí, je potřeba si také uvědomit, jak stroje ovlivňují naše chování a náš život. Zda se nedostáváme do situace, která je mnohem nebezpečnější? A zda se při používání těchto systémů nestávají stroje z nás? Zda neupouštíme od toho, co jsme se naučili a jaké znalosti jsme získali? Zda se nezpronevřujeme svým hodnotám, představitosti a intuici? Tyto otázky mohou být dalším podnětem pro budoucí rozšíření tohoto tématu, které půjde hlouběji či bude zohledňovat další oblasti, jež etika prostupuje, jako jsou práce, hodnotový systém, vzdělání ...

## Seznam literatury

### Primární zdroje:

- ANDERSON, R., *Why Information Security is Hard—An Economic Perspective*. In Proc. 17th Annual Computer Security Applications Conference, 2001.
- ANOM, B.Y. „Ethics of Big Data and artificial intelligence in medicine.“ In: *Ethics, Medicine and Public Health*. [online]. 2020, roč. 15. 10. [cit. 2024-04-01]. Dostupné z: <https://doi.org/10.1016/j.jemep.2020.100568>
- ARNOLD, John E. (1959a) 2016. *Creative Engineering: Promoting Innovation by Thinking Differently*. Edited with an introduction and biographical essay by William J. Clancey. Stanford Digital Repository. Retrieved 13 December 2016.
- JEDLIČKOVÁ, Anetta. „Etické aspekty rozvoje umělé inteligence.“ In: *Anthropologia integra: časopis pro obecnou antropologii a příbuzné obory*. [online]. Brno: Masarykova univerzita, 2010-, 13/2022/2, str.57 [cit.2024-03-07]. Dostupné z: [https://journals.muni.cz/anthropologia\\_integra/article/view/32983/28375](https://journals.muni.cz/anthropologia_integra/article/view/32983/28375)
- ARISTOTELÉS. *Etika Nikomachova*. Páté nezměněné vydání. Praha: Rezek, 2021.
- ASIMOV, Isaac. *Já, robot*. Přeložil Oldřich ČERNÝ, přeložil Alexandr KRAMER, přeložil Zuzana MEYEROVÁ. Trifid (Triton). Praha: Triton, c2012.
- BRYNJOLFSSON, Erik. *Druhý věk strojů: práce, pokrok a prosperita v éře špičkových technologií*. Přeložil Filip DRLÍK. V Brně: Jan Melvil Publishing, 2015. Pod povrchem.
- DENNETT, D.C. *Consciousness Explained*. Boston: Brown and Co., 1991a.
- ENGHOLM, Ida. *Quick guide to design thinking*. 1. vyd. Copenhagen.
- FERRARI, A., BREČKO, B.N., PUNIE, Y. „DIGCOMP: A Framework for Developing and Understanding Digital Competence in Europe.“ 2014. Dostupné z: [https://www.researchgate.net/publication/282860020\\_DIGCOMP\\_a\\_Framework\\_for\\_Developing\\_and\\_Understanding\\_Digital\\_Competence\\_in\\_Europe/link/561fbcf208aed8dd19402c64/download?\\_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19](https://www.researchgate.net/publication/282860020_DIGCOMP_a_Framework_for_Developing_and_Understanding_Digital_Competence_in_Europe/link/561fbcf208aed8dd19402c64/download?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19)
- FLORIDI, L., COWLS, J., BELTRAMETTI, M., CHATILA, R., CHAZERAND, P., DIGNUM, V., LUETGE, Ch., MADELIN, R., PAGALLO, U., ROSSI, F., SCHAFFER, B., VALCKE, P., VAYENA, E. „AI4People-an Ethical Framework for a Good AI Society: Opportunities, Risks, Principles and Recommendations,“ In: *Minds and Machines*, 2018, 28,č.4.
- GARDNER, Howard. *Frames of mind: the theory of multiple intelligences*. New York: BasicBooks, 1983. xxxii, 440.
- GUNKEL J. DAVID. „Perspectives on ethics of AI“ In *The Oxford Handbook of Ethics of AI*, edited by DUBBER, D. Markus, PASQUALE, Frank, DAS, Sunit. Oxford University Press, 2020.
- HALL, J. Storrs. „Ethics for Machines,“ *KurzweilAI.net*. [online]. [cit. 2001-07-05]. Dostupné z <http://www.kurzweilai.net/ethics-for-machines>
- HEIDEGGER, Martin a PETŘÍČEK, Miroslav. *Bytí a čas*. Třetí, opravené vydání. Přeložil Ivan CHVATÍK, přeložil Pavel KOUBA, přeložil Jiří NĚMEC. Knihovna novověké tradice a současnosti. Praha: OIKOYMENH, 2018.
- HERWEIJER, C. „8 Ways AI Can Help Save the Planet.“ In: *World economic forum*. [online]. 13.8.2023]. Dostupné z: <https://www.weforum.org/agenda/2018/01/8-ways-ai-can-help-save-the-planet/>
- HUME, D. *A Treatise of Human Nature*. (2.vydání). Selby-Bigge, Los Angeles, Oxford: Clarendon,1978.

- JANKOVSKÝ, Jiří. *Etika pro pomáhající profese*. Praha: Triton, 2003.
- KANT, Immanuel. *Základy metafyziky mravů*. 2. vyd. Filozofické dědictví. Praha: Svoboda, 1990.
- KREMPASKÝ, Július, Ivan. *Synergetika*. Vydavatelství STU Bratislava, 1994.
- HAVEL I. M. „Přirozené a umělé myšlení jako filozofický problém.“ In *Umělá inteligence*. MAŘÍK, Vladimír, Olga ŠTĚPÁNKOVÁ a Jiří LAŽANSKÝ a kolektiv. Praha: Academia, 1993-.
- „Preliminary Statement of Policy Concerning Automated Vehicles“ In *National Highway Traffic Safety Administration*, [online]. 2013. s. 4. [cit. 2023/04/12]. Dostupné z [https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/automated\\_vehicles\\_policy.pdf](https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/automated_vehicles_policy.pdf)
- NEČASOVÁ, Mirka. *Úvod do filozofie a etiky v sociální práci*, Brno: Masarykova univerzita. 2001.
- SEARLE, John R. *Mysl, mozek a věda*. Přeložil Marek NEKULA. Praha: Mladá fronta, 1994. Váhy (Mladá fronta).
- SCHWARTZ, R., DODGE, J., SMITH, N. A., & ETZIONI, O. „Green ai.“ In: *Communications of the ACM*, 2020. 63(12).
- SOLOVE, D. J. „A Taxonomy of Privacy.“ In: *University of Pennsylvania Law Review*, [online]. 2006,154(3). [cit.2023-07-08].
- STEFFEN, W. & CRUTZEN, PAUL & MCNEILL, JOHN. „The Anthropocene: Are Humans Now Overwhelming the Great Forces of Nature.“ In: *AMBIO A Journal of the Human Environment*. (2008) v.36.
- ŠIMEK, V. „Morální odpovědnost.“ In *FIZAMI* [online]. [cit. 2017-10-12]. Dostupné z <https://www.fizami.cz/video/moralni-odpovednost/>
- THOMPSON, Mel. *Přehled etiky*. Vyd. 1. Praha: Portál, 2004.
- VARIAN, H. *Intermediate Microeconomics - A Modern Approach*. páté vydání, WW Norton and Company, New York, 1999.
- Všeobecná deklarace lidských práv*. B.m.: Společnost pro lidská práva, 1968.

#### Sekundární zdroje:

- ANZENBACHER, Arno. *Křesťanská sociální etika: úvod a principy*. 2. vydání. Přeložil Karel ŠPRUNK. Brno: CDK, Centrum pro studium demokracie a kultury, 2015.
- BAŠTECKÁ, Bohumila a Petr GOLDMANN, 2001. *Základy klinické psychologie*. Vyd. 1. Praha: Portál.
- BOVENS, Luc. „The Ethics of Nudge.“ In: Mats J. Hansson & Till Grüne-Yanoff (eds.), *Preference Change: Approaches from Philosophy, Economics and Psychology*. Berlin: Springer, Theory and Decision Library A. Chapter 10. 2008.
- BRADNOVÁ, Hana. *Encyklopedický slovník*. Praha: Odeon, 1993. Klub čtenářů (Odeon).
- BROWNLEE, Jason. „A Gentle Introduction to Generative Adversarial Networks (GANs). Machine Learning Mastery.“ In: *Generative Adversarial Network* [online]. 2019. [cit. 2024-02-11]. Dostupné z: <https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/>
- BRYSON J. Joanna. „The artificial intelligence of the ethics of artificial intelligence. An Introductory Overview for Law and Regulation“ In *The Oxford Handbook of Ethics of AI*, edited by DUBBER, D. Markus, PASQUALE, Frank, DAS, Sunit. Oxford University Press, 2020.

- CARR, Nicholas G. *Skleněná klec: automatizace a my*. Přeložil Štěpán KAŇA. Brno: Emitos, 2015.
- CARR, Nicholas G. *Nebezpečná mělčina: jak internet mění náš mozek: analýza stavu lidské psychiky v době digitální*. Přeložil Jaroslava PŘEROVSKÁ. V Praze: Dauphin, 2017.
- COECKELBERGH, Mark. *Etika umělé inteligence*. Filosofia, nakladatelství Filozofického ústavu AV ČR, v.v.i.,2023.
- DIGNUM, Virginia. "Responsibility and artificial intelligence" In *The Oxford Handbook of Ethics of AI*, edited by DUBBER, D. Markus, PASQUALE, Frank, DAS, Sunit. Oxford University Press, 2020.
- DREYFUS, HUBERT and KELLY, SEAN D. *All Things Shining: Reading the Western Classics to Find Meaning in a Secular Age*. New York, 2011. NY: Free Press
- DUBBER, D. Markus, PASQUALE, Frank, DAS, Sunit. *The Oxford Handbook of Ethics of AI*. Oxford University Press, 2020.
- SIMON, Judith. „Distributed Epistemic Responsibility in a Hyperconnected Era.“ In: *The Onlife Manifesto: Being Human in a Hyperconnected Era*. Edited by FLORIDI, Luciano. Imprint: Springer, 2015.
- FRISCHMANN, B., & SELINGER, E. *Re-Engineering Humanity*. Cambridge: Cambridge University Press,2018.
- FRY, Hannah. *Hello world: jak zůstat člověkem ve světě algoritmů*. Přeložil Roman ŠOLC. V Praze: Vyšehrad, 2020.
- GEBRU TIMNIT. "Race and gender" In *The Oxford Handbook of Ethics of AI*, edited by DUBBER, D. Markus, PASQUALE, Frank, DAS, Sunit. Oxford University Press, 2020.
- CHAMBERS, Amy. „There’s a Reason Siri, Alexa and AI Are Imagined as Female – Sexism“ In *The Conversation* [online]. 2018. Dostupné z: <http://theconversation.com/theres-a-reason-siri-alexa-and-ai-are-imagined-as-female-sexism-96430/>.
- MARR. Bernard. „27 Incredible Examples of AI and Machine Learning in Practice,“ In: *Forbes*. [online]. 30.4.2018. [cit. 2023-04-01]. Dostupné z: <https://www.forbes.com/sites/bernardmarr/2018/04/30/27-incredible-examples-of-ai-and-machine-learning-in-practice/#6b37edf27502>
- OpenAI. (2023). GPT-4 [ChatGPT model]. OpenAI.
- PEREGRIN, Jaroslav. *Kapitoly z analytické filosofie*. Praha: Filosofia, 2005.
- RUISEL, Imrich. *Základy psychologie inteligence*. Praha: Portál, 2000.
- RUSANEN, Anna-Mari. „Chapter 1: What is AI ethics?“ In: *Ethics of ai* [online]. [cit. 2022-06-10]. Dostupné z: <https://ethics-of-ai.mooc.fi/chapter-1/1-a-guide-to-ai-ethics>
- RUSANEN, Anna-Mari. „Chapter 2: What should we do?“ In: *Ethics of ai* [online]. [cit. 2022-06-10]. Dostupné z: <https://ethics-of-ai.mooc.fi/chapter-2/1-what-should-we-do>
- RUSANEN, Anna-Mari. „Chapter 3: Who should be blamed?“ In: *Ethics of ai* [online]. [cit. 2022-06-10]. Dostupné z: <https://ethics-of-ai.mooc.fi/chapter-3/1-algorithms-and-accountability>
- RUSANEN, Anna-Mari. „Chapter 4: Should we know how AI works?“ In: *Ethics of ai* [online]. [cit. 2022-06-10]. Dostupné z: <https://ethics-of-ai.mooc.fi/chapter-4/1-transparency-in-ai>
- RUSANEN, Anna-Mari. „Chapter 6: Should AI be fair and non-discriminative?“ In: *Ethics of ai* [online]. [cit. 2022-06-10]. Dostupné z: <https://ethics-of-ai.mooc.fi/chapter-6/1-what-is-fairness>

- TEGMARK, Max. *Život 3.0: člověk v éře umělé inteligence*. Přeložil Markéta IVÁNKOVÁ. Praha: Argo, 2020. Zip (Argo: Dokořán).
- TVRDÝ, Filip. *Turingův test, Filosofické aspekty umělé inteligence*. Olomouc, 2011. Disertační práce (PhD.). Univerzita Palackého v Olomouci, Filosofická fakulta, Katedra filosofie.
- Učebnice obecné neurologie*. Brno: Masarykova univerzita, 2003.
- ZELINKA, Ivan. *Umělá inteligence: hrozba nebo naděje?* Praha: BEN-technická literatura, 2003.