

UNIVERZITA KARLOVA



FAKULTA HUMANITNÍCH STUDIÍ

obor sociální a kulturní ekologie

Dominika Paukertová, BSc.

UMĚLÁ INTELIGENCE VE SVĚTLE SOCIÁLNÍ VĚDY

Artificial intelligence in the light of social science

Diplomová práce

Vedoucí práce: PhDr. Ivan Rynda

Praha 2022



## **Prohlášení**

Prohlašuji, že jsem předkládanou práci zpracovala samostatně a použila pouze uvedené prameny a literaturu. Práce nebyla využita k získání jiného nebo stejného titulu. Současně dávám svolení k tomu, aby tato práce byla zpřístupněna v příslušné knihovně UK a v elektronické databázi vysokoškolských kvalifikačních prací a v souladu s autorským právem používána ke studijním účelům.

V Praze dne 30. 06. 2022

Dominika Paukertová BSc.

.....

## **Poděkování**

Ráda bych poděkovala svému vedoucímu diplomové práce PhDr. Ivanu Ryndovi za cenné rady a vstřícnost při vedení mé práce. Dále bych chtěla poděkovat Ing. Vojtěchu Nezdarovi za odborné rady z oblasti informatiky, které mi umožnily hlubší vhled do problematiky.

V Praze dne 30. 06. 2022

## **Abstrakt**

Předložená diplomová práce se zabývá fenoménem umělé inteligence, na který se dívá pohledem společenské vědy. Konkrétně se snaží odpovědět na otázku, zda dokázali společenské vědy anticipovat budoucnost v případě umělé inteligence. Pro zodpovězení této otázky je v následující práci použita kombinace metod symptomálního čtení a hermeneutiky. Ty spočívají ve čtení zvolené literatury za účelem nalezení hlubšího významu, který se na první pohled mohl zdát skrytý. Pro analýzu jsem jako reprezentanty společenských věd zvolila Ulricha Becka, Anthonyho Giddense a Niklase Luhmanna, jelikož každý z nich se zabývá vztahem společnosti a přírody. Lze tak předpokládat jejich holistický pohled na svět, který by měl obsáhnout i téma umělé inteligence. Práce nejprve mapuje samotnou umělou inteligenci, její vývoj, druhy a hrozby. Závěrem první části jsou definovány tři vybrané oblasti aktuálních rizik a to: nejasnost co vlastně AI je a není a nejasné představy lidí o ní, možnost manipulace textu a obrazu a konečně problém zastaralých dat a předpojatosti kódu. Před samotnou analýzou jen krátce přiblížuji několik pojmů ze společenskovední literatury pro lepší orientaci v samotné analýze. Tyto pojmy jsou: riziko, důvěra a budoucnost. V samotné analýze se za použití zvolených metod snažím zkoumat problematiku umělé inteligence pohledem společenské vědy v kontextu zvolených problematických oblastí AI a nacházet skryté souvislosti a významy. V rámci analýzy jsem zjistila, že ke každému ze zvolených rizik AI se vyjádřil alespoň jeden autor a tím si kladně odpověděla na otázku, zda společenská věda dokázala anticipovat budoucnost umělé inteligence. To potvrzuje důležitost partnerství společenských a přírodních věd v rámci řešení otázek nejen doby přítomné, ale i budoucí.

## **Klíčová slova**

umělá inteligence, riziko, důvěra, riziková společnost, společenské vědy

## **Abstract**

This diploma thesis deals with the phenomenon of artificial intelligence, which is viewed through the lens of social science. Specifically, it seeks to answer the question of whether social science has been able to anticipate the future in the case of artificial intelligence. In order to answer this question, a combination of the methods of symptomatic reading and hermeneutics is used in the following paper. These consist of reading the selected literature in order to find deeper meanings that may have seemed hidden at first glance. For the analysis I have chosen Ulrich Beck, Anthony Giddens and Niklas Luhmann as representatives of the social sciences, as each of them deals with the relationship between society and nature and thus their holistic worldview can be assumed to include the topic of artificial intelligence. The thesis first deals with artificial intelligence itself, its development, types and threats. It concludes the first part by defining three selected areas of current risks, namely: the ambiguity of what AI actually is and is not and people's unclear ideas about it, the possibility of text and image manipulation, and finally the problem of outdated data and biased code. Before the analysis itself, I briefly introduce a few concepts from the social science literature for better orientation in the analysis. These concepts are: risk, trust and future. In the analysis, using the chosen methods, I try to explore the issues of artificial intelligence through social science point of view in the context of the chosen AI problem areas and find hidden connections and meanings. In the analysis, I found that at least one author dealt with each of selected AI risks, thereby answering positively the question of whether social science has been able to anticipate the future of AI. This also confirms the importance of the partnership between the social sciences and the natural sciences in addressing issues not only of the present but also of the future.

## **Key words**

artificial intelligence, risk, trust, risk society, social sciences

## Obsah

1	Úvod.....	9
1	Umělá inteligence .....	11
1.1	Definice AI.....	11
1.2	Typy AI.....	13
1.2.1	Úzká (slabá) umělá inteligence (ANI) .....	13
1.2.2	Obecná (silná) umělá inteligence (AGI) .....	14
1.2.3	Superinteligence (ASI).....	14
1.3	Přístupy k AI .....	15
1.3.1	Symbolická AI (Good Old-Fashioned AI).....	15
1.3.2	Konekcionismus .....	16
1.4	Testování AI.....	17
1.4.1	Turingův test .....	17
1.4.2	Argument čínského pokoje.....	17
1.4.3	Winograd Schema Challenge (WSG) .....	18
1.4.4	Lovelace test.....	19
1.5	Historie vývoje umělé inteligence.....	20
1.5.1	Nové pojetí stroje a vznik oboru .....	20
1.5.2	První AI Winter .....	22
1.5.3	Vzestup expertních systémů.....	23
1.5.4	Ekonomická bublina a druhá AI Winter .....	25
1.5.5	Devadesátá léta, neuronové sítě a Gary Kasparov .....	26
1.5.6	Autonomní vozidla a hluboké učení.....	28
1.5.7	Watson, Tesla, AlphaGo a pokroky ve strojovém učení.....	30
1.6	Technologie AI .....	34
1.6.1	Expertní systémy .....	34
1.6.2	Machine learning (strojové učení).....	35
1.6.3	Znalostní grafy .....	38
1.7	Současnost.....	39
1.7.1	Zpracování textu.....	39
1.7.2	Zpracování obrazu.....	40
1.7.3	Robotika a mobilita .....	41
1.7.4	Zdravotnictví.....	42
1.7.5	Finance .....	43

1.7.6	Cesta k obecné inteligenci.....	44
1.7.7	Shrnutí.....	45
2	Metodologie .....	46
2.1	Althusserovo symptomální čtení.....	46
2.2	Hermeneutika .....	47
2.3	Výběr literatury k analýze.....	48
3	Stručný úvod k analýze .....	50
3.1	Riziko .....	50
3.2	Důvěra .....	53
3.3	Budoucnost.....	56
4	Analytická část.....	60
4.1	Věda .....	60
4.1.1	Věda v rizikové společnosti .....	60
4.1.2	Pozice laika v expertních systémech.....	66
4.1.3	Důvěra ve vědění.....	68
4.2	Technika.....	69
4.2.1	Technika, komunikace a Niklas Luhmann.....	69
4.2.2	Technika bez člověka .....	73
4.2.3	Člověk není proměnná .....	74
5	Diskuze.....	79
5.1	Výzkumné otázky .....	79
5.2	Výběr autorů .....	85
6	Závěr .....	87
7	Seznam použité literatury.....	89
7.1	Metodologická literatura .....	89
7.2	Literatura .....	89
8	Přílohy.....	99



# 1 Úvod

*„Umělá inteligence je v současnosti velice oblíbeným a žhavým tématem. Její možnosti jsou mnohé a její využití je často omezeno pouze naší fantazií. Umělá inteligence je však také kontroverzní téma a mnoho lidí ji vnímá s obavami.*

*Právě proto byla zvolena tato diplomová práce. Cílem této práce je představit umělou inteligenci a její možnosti včetně výhod a nevýhod. Dále se práce zaměří na obavy spojené s umělou inteligencí a na to, zda jsou oprávněné, nebo ne.*

*Práce je rozdělena na několik částí. V první části je představena umělá inteligence jako taková, její vývoj a možnosti. Dále jsou zde uvedeny obavy spojené s umělou inteligencí, které jsou rozděleny do několika kategorií. V závěrečné části je pak shrnuto, co bylo v práci uvedeno a jsou zde uvedeny závěry autora.“*

Poznali jste, že úvod výše vygenerovala umělá inteligence<sup>1</sup>? To zvládla jenom na základě textového příkazu „napiš úvod diplomové práce o umělé inteligenci“. Na tomto jednoduchém příkladu by se dal vystihnout problém, kterým se chci v následující diplomové práci zabývat. Umělá inteligence, nebo také zkráceně AI z anglického názvu Artificial Intelligence, v lidech budí emoce ať se v jejím fungování vyznají, nebo si ji spojují jen se žánrem sci-fi. Teprve minulý rok (2001) v březnu byl ten stejný model AI předmětem populárního článku o umělé inteligenci, která tvořila balící hlášky, znějící spíše komicky. Nyní o pouhý rok později vytvořila úvod k diplomové práci, který by možná prošel bez povšimnutí. Je tedy zřejmé, jak rychle se tato technologie vyvíjí a zlepšuje.

Ačkoli je pro mnohé představa AI velice abstraktní, je již delší dobu součástí světa a životů nás všech. Rozpoznává otisk prstu při přihlášení do mobilního telefonu, zlepšuje vyhledávání na internetu a pokud jste někdy zadávali zdánlivě nesmyslný krátký text nebo poznávali na čtvercích obrázek autobusu tak věřte, že jste se podíleli na tréninku modelu umělé inteligence. AI je ale také integrována v systémech řízení dopravy, kontrole kvality vody, leteckém průmyslu, zdravotnictví a výčet by mohl pokračovat dál. Pokud ve vás tato představa budí nejistotu či strach, věřte, že nejste sami. AI je nejen průlomovou technologií zjednodušující život a zkvalitňující procesy, ale i v lidech budí obavy často oprávněně.

---

<sup>1</sup> Konkrétně jazykový model GPT-3, využívající hlubokého učení k vytváření textu.

Viz <https://beta.openai.com/playground>

Všemi těmito aspekty, jak benefity, ale hlavně negativy, riziky a místem člověka ve světě umělé inteligence se budu v následující práci zabývat.

Na to, že už jenom název „umělá inteligence“ napovídá spojení světa technického a sociálního je toto téma málo reflektované oběma pohledy, kdy by ani jeden výrazně nepřevyšoval, a zvláště pak v českých kruzích. Proto jsem rozhodla tento obor informatiky, součásti přírodních věd, podrobit drobnohledu společenské vědy zastoupené několika vybranými autory. Cílem této práce je zjistit, zda společenská věda dokázala anticipovat rychle se měnící technologii umělé inteligence dříve, než se jí dostalo podoby, jakou má dnes. To nám poskytne představu i o výpovědní hodnotě společenských věd pro oblasti přírodních věd.

Jak jsem zmínila, jednou z mých motivací při práci na tomto textu je i nestrannost vůči oběma oblastem. Proto bych v teoretické části ráda nabídla ucelený pohled na umělou inteligenci, její historii a současné možnosti a hrozby. I díky tomu snad bude možné udělat si lepší představu o AI a částečně ji zbavit i puncu tajemna. Další bude metodologická část, ve které se budu zabývat postupem, jak k nastavenému cíli diplomové práce dospěji. Následovat bude krátký úvod k analýze, který osvětlí základní pojmy ze společenských věd, důležitých pro lepší orientaci v analýze. Ta pak nabídne rozbor děl vybraných autorů a hledání podobností a rozdílností s literaturou zabývající se umělou inteligencí. Zjištěními práce se budu zabývat v části diskuze a výsledky budou shrnuty v závěru.

V příloze naleznete Obrázek 1, který ukazuje časovou osu vývoje milníků společenské vědy s AI.

# 1 Umělá inteligence

Umělá inteligence je v současnosti velice aktuálním a populárním tématem. Její možnosti jsou rozličné a využití je často omezeno pouze naší fantazií. Umělá inteligence je také spjata s kontroverzemi emocemi a mnoho lidí ji vnímá s obavami. A i to byl jeden z důvodů, proč jsem si zvolila ke zpracování toto téma. V této teoretické části chci představit umělou inteligenci a její reálné možnosti včetně výhod a nevýhod. Také se chci zaměřit na obavy z ní a jak moc reálné se jeví. Mým cílem bylo představit tuto technologii tak detailně, aby o jejích rizicích bylo možné uvažovat v reálných možnostech a věnovat se tak aktuálním hrozbám v analytické části.

Práce je rozdělena na několik oddílů. V první části je představena umělá inteligence jako taková, její vývoj a možnosti. Dále jsou zde uvedeny obavy spojené s umělou inteligencí, které jsou rozděleny do několika kategorií. V závěrečné části je pak shrnuto, co bylo v práci uvedeno a jsou zde uvedeny závěry autora.

## 1.1 Definice AI

*„Dostatečně pokročilou technologii nelze odlišit od magie.“*

Arthur Charles Clarke

Odpověď na otázku “Co je to umělá inteligence?” velmi záleží na tom, kdo je dotazován. Laik si pravděpodobně představí nějakého robota, technicky znalejší jedinec třeba hlasového asistenta či doporučený obsah na webu a výzkumník z technické oblastí odpoví, že je to algoritmů se schopností dobrat se výsledku bez jasných instrukcí.

V dnešním světě je téměř nemožné vyhnout se některému z využití umělé inteligence, ať už při používání chytrého telefonu, prohlížení webu nebo při pracovním procesu. Odemykání různých zařízení pomocí obličejů se stalo téměř standardem a téměř každý z nás na nějaké platformě získává personalizovaný obsah “na míru”. Bez nadsázky lze říci, že algoritmy ovlivňují spoustu aspektů našeho každodenního života, aniž bychom si uvědomovali, v jaké míře a jak přesně.

Sci-fi literatura a filmy operují s tímto pojmem již několik desítek let a mohlo by se tak zdát, že jde o něco ustáleného. Opak je však pravdou. Vymezení pojmu AI je obtížnější, než by se na první pohled mohlo zdát, a to hned z několika důvodů. Většina pojmu v informatice a jiných technických vědách je jasně ukotvených a neměnných v čase. Některé technologie a postupy, které byli před lety považovány za AI, jsou dnes již jen pokročilá statistika vyučovaná na střední škole.

Nick Bostrom v roce 2003 zdůraznil, že cíl umělé inteligence je nezávislý na její inteligenci (definované jako její schopnost dosáhnout jakéhokoli cíle). Jediným cílem šachového počítače může být vítězství v šachu. Existují však i turnaje v takzvaných prohraných šachách, kde je cílem přesný opak. Počítače, které v nich soutěží, nejsou méně inteligentní než ty naprogramované na vítězství. [Tegmark, 2017: 238] Britský informatik Stuart Russell definuje AI jako “stroje jsou inteligentní do té míry, v níž se dá očekávat, že pomocí svého konání dosáhnou svých záměrů.[...] a která vychází z obecné definice konceptu lidské inteligence.“ Sám autor následně tuto definice považuje za nevhodnou, jelikož nezohledňuje prospěšnost a AI by měla spíše sledovat naše záměry, nežli svoje. [Russel, 2019: 15-16]

O pevně danou definici se již delší dobu pokouší i Evropská Unie. Cílem této iniciativy je zejména přiblížení této oblasti i laické veřejnosti. Lepší porozumění by mělo přispět k větší věrohodnosti a důvěře v AI. V jednom z posledních návrhů z roku 2021 je AI definována jako “software, který je vyvinut pomocí jedné nebo více technik a přístupů uvedených v příloze I<sup>2</sup>, a který může pro danou sadu cílů definovaných člověkem generovat výstupy, jako je například obsah, predikce, doporučení nebo rozhodnutí ovlivňující prostředí, s nimiž komunikují”. [Návrh Nařízení EP a Rady, 2021] Tato definice způsobuje debaty ve vědecké komunitě kvůli poměrně vágnímu a obecnému přístupu ve zmíněné příloze. Jsou zde zmíněné běžné statistické metody, jako Bayesův odhad, jejichž použití nemusí mít nic společného s umělou inteligencí. Problém je to zejména z legislativních důvodů, jelikož EU pracuje na pravidlech pro regulaci. [Carpenter, 2022]

Jeden ze zakladatelů vědního oboru AI John McCarthy údajně s oblibou říkal “Jakmile to bude fungovat, nikdo tomu už nebude říkat umělá inteligence.”. [Meyer, 2011] Tento jev je některými označován jako “efekt AI” - jakmile technologii porozumíme, již nejde o skutečnou inteligenci a veškeré kouzlo zmizí. Kevin Kelly, dlouholetý redaktor časopisu Wired, ve svém článku napsal: “Dříve bychom řekli, že pouze superinteligentní umělá inteligence dokáže řídit auto, porazit člověka v Jeopardy! nebo v šachu. Jakmile však umělá inteligence každou z těchto věcí zvládla, považovali jsme tento úspěch za zjevně mechanický a stěží hodný označení skutečná inteligence. Každý úspěch v oblasti umělé inteligence ji nově definuje.” [Kelly, 2014]

---

<sup>2</sup> Příloha 1 příslušného nařízení. Není součástí této práce.

Osobně nejvíce souhlasím s definicí používanou Univerzitou v Louisianě v Lafayette: “Umělá inteligence je studium člověkem vytvořených výpočetních zařízení a systémů, které lze přimět k tomu, aby se chovaly způsobem, který bychom mohli označit za inteligentní.” [Berkeley, 1997]. Tato definice vystihuje povahu oboru a zohledňuje možný rozdílný pohled na něj.

## 1.2 Typy AI

### 1.2.1 Úzká (slabá) umělá inteligence (ANI)

Úzká, někdy označovaná jako *slabá umělá inteligence* je definována podobně problematicky jako umělá inteligence. Tedy není zde jednoznačný konsensus a definice mohou být i zavádějící. Pro účel této práce jsem se rozhodla čerpat z několika zdrojů a spíše popsat, co lze považovat za *úzkou ANI*<sup>3</sup>, nežli se pokoušet o přesnou definici. [Coeckelbergh, 2020] [Wooldridge, 2021] [Tegmark, 2017]

Na úvod je vhodné říci, že veškeré současné aplikace (veřejně známé) spadají do této kategorie. Dokáží provádět úzce definovaný úkol lépe než člověk. Tento typ umělé inteligence nemá vědomí, není vnímavý a neřídí se emocemi. Pracuje v reálném čase a v předem definovaném rámci. Tento rámec není schopen překročit. Některé současné aplikace se nám mohou zdát sofistikovanější než-li ve skutečnosti jsou - například hlasový asistenti kteří s námi komunikují a zpracovávají tak přirozený jazyk nebo řídí autonomní vozidla.

Ani tyto asistenti však nedokáží myslet sami za sebe a jen používají předdefinovanou sadu řešení určitých problémů. Toto lze definovat na dvou příkladech použití/předpovědi počasí a osobní otázce, například o smyslu života. Při předpovědi počasí asistent použije předdefinovanou akci a v drtivé většině případů podá velice kvalitní odpověď. V případě osobnější otázky jen zpracuje hlasový povel a zadá ho do vyhledávače, takže odpověď je vágní nebo nedává příliš smysl.

Dalším příkladem, kde se možná překvapivě jedná o úzkou inteligenci, mohou být autonomní vozidla. Ačkoliv vozidla monitorují okolí a reagují na něj, nikdy to není řízeno emocemi jako u lidských řidičů. Na druhou stranu nemůže ani být tak flexibilní a plynulá jako člověk. Autonomní vozidlo se skládá z více systémů úzké umělé inteligence, které spolu navzájem komunikují, ale nelze ho považovat za něco více sofistikovaného.

---

<sup>3</sup> V angličtině jako narrow nebo weak ANI

Ačkoliv tento typ inteligence nazýváme poněkud “nelichotivě” jako inteligenci slabou, stále se jedná o velmi důležitý objev a ukázkou lidského pokroku. Systémy úzké umělé inteligence dokáží plnit úkoly podstatně rychleji než i ty nejzkušenější pracovníci a zpracovat opravdu velké množství dat. To vede k větší produktivitě a efektivitě ve spoustě odvětví. Michael Wooldridge navíc poznamenává, že v AI komunitě se toto označení používá jen zřídka a aplikace úzké AI jsou prostě “aplikace AI” [Wooldridge, 2021: 30]. Stejný trend můžeme zaznamenat i ve veřejném prostoru.

### 1.2.2 Obecná (silná) umělá inteligence (AGI)

Dalším logickým krokem po vytvoření úzké umělé inteligence je inteligence obecná. Taková inteligence by myslela stejným způsobem jako lidský mozek a měla by mít potenciál ho překonat [Bostrom, 2014]. Aby se opravdu takový systém umělé inteligence podobal skutečné lidské inteligenci měl by mít stejné kognitivní funkce včetně vnímání, jednání, vědomí, záměrů a pravděpodobně i emocí [Kurzweil, 2006].

Ve vědeckých kruzích se vedou diskuze, zda je tato obecná inteligence vůbec dosažitelná [Coeckelbergh, 2020]. Někteří tuto možnost přímo odmítají a někteří zastávají názor, že v blízké době rozhodně nepřijde [Boden, 2016]. Jedním z prvních a také nejznámějších kritiků je John Searle, který si vytvořil argument čínského pokoje<sup>4</sup> a dospěl k závěru že obecná inteligence není možná. Naopak Ray Kurzweil ji předpovídá na rok 2029. [Kurzweil, 2006: 196]

### 1.2.3 Superinteligence (ASI)

*“Jsme jako děti, které si hrají s bombou”*

*Nick Bostrom*

Superinteligence je teoretický vývojový stupeň obecné inteligence. Bostrom ji definuje jako: „Jakýkoliv intelekt, který lidské kognitivní výkony dalece překonává prakticky ve všech relevantních oblastech.“. Podle Bostroma [Bostrom, 2014] by naše situace byla srovnatelná s gorilami. Jejich osud je dnes zcela závislý na nás.

Bostrom definuje několik cest, jak by mohla ASI vzniknout. Jedna z nich je díky tzv. explozi inteligence. AI by byla schopna vytvořit *rekurzivní sebezdokonalování*, tedy vytvářet vylepšené verze sama sebe, které by opět byly schopné vytvořit lepší verzi a tak dále. Další

---

<sup>4</sup> Kapitola Testování AI

možnou cestou je emulace či dokonce nahrávání celého mozku: biologický by mohl být vymodelován, naskenován a reprodukován v rámci softwaru [Bostrom, 2014].

Max Tegmark si představuje vizi umělé inteligence, která se stane všemocnou a bude tak řídit celou planetu. Yuval Harari ve své knize píše o světě, v němž už lidé nevládnou, ale pouze uctívají data a svěřují svá rozhodnutí algoritmům. Po zničení všech humanistických iluzí a liberálních institucí sní lidé o splnutí s tokem dat. *“Algoritmus může být zpočátku vyvinut člověkem, ale jak roste, jde svou vlastní cestou a jde tam, kam ještě žádný člověk nešel - a kam ho žádný člověk nemůže následovat.”* [Coeckelbergh, 2020: 17].

### 1.3 Přístupy k AI

Již od počátku se výzkum AI dělil na dva odlišné proudy – symbolickou AI a konekcionismus (někdy označován jako neuronové sítě). Hlavním rozdílem byl přístup k vytvoření umělé inteligence a to, zda použít analogii počítače nebo analogii mozku.

#### 1.3.1 Symbolická AI (Good Old-Fashioned AI)

Symbolická umělá inteligence je nejstarším přístupem, který v rámci AI rozlišuje me. Tento přístup byl dominantním paradigmatem od poválečného období až do pozdních 80. let 20. století. Často se v jeho případě můžeme setkat s označením “good old-fashioned artificial intelligence” (“stará dobrá umělá inteligence”, GOFAI), poprvé použitým Johnem Hauglenadem v jeho knize *“Artificial Intelligence: The Very Idea”*, která zkoumala filozofické důsledky výzkumu umělé inteligence [Nilsson, 2009: 391].

Přístup klade důraz na popis inteligence pomocí lidsky čitelných symbolů, které reprezentují entity z reálného světa a matematickou logiku, k vytvoření pravidel pro možnost jejich manipulaci. Vychází hlavně z toho, že počítače jsou jen manipulátory se symboly, přičemž symboly jsou jen binární (jedničky a nuly) a jsou uspořádány do slov, řádků a stránek. Počítač s těmito symboly umí manipulovat (přesouvat, mazat, zapisovat), porovnávat je a vyhledávat určité vzory. Všechno složitější je jen kombinace těch jednoduchých úloh.

Jedním z hlavních rozdílů mezi symbolickou AI a strojovým učením je místo kde k tomuto učení dochází. Při strojovém učení se algoritmus nejčastěji učí pravidla během zjišťování korelací mezi vstupy a výstupy. U symbolického přístupu se příslušný expert musí prvně naučit pravidla, podle kterých spolu dva jevy souvisí a následně je zanést do programu. Zadaná pravidla značně omezují možnosti, jak se může program dále učit a zlepšovat (prakticky tuto možnost eliminuje).

Největším problémem symbolické AI je možnost revize již zadaných pravidel. Čím více pravidel je zadáno, tím je systém “chytřejší” - zároveň však platí, že se změnou jednoho pravidla dochází ke změně všech navázaných pravidel. Dalším problémem je kontext – program sám většinou neví, co symboly znamenají a jakou mají spojitost s “okolním světem”. Posledním zásadnějším problémem je v samotném zadávání pravidel. Pokud je zadáno špatně, zpravidla systém vrátí nesmyslný výsledek. [Nilsson, 2009]

Logicky se u těchto systémů nabízí otázka pro koho jsou symboly dobré. Program nepotřebuje pracovat s přirozenou řečí a jistě by efektivněji operoval pomocí vektorů či jiných technických dialektů. V nejlepším případě by náš přirozený jazyk měl být jen rozhraním, kterým k programu přistupujeme a měli bychom umělou inteligenci nechat být skutečně inteligentní. Bez přirozeného jazyka by však bylo těžké dávat najevo, jak jsme chytří a jen by to rozšířilo strach, že nás AI jednou zničí, jelikož by bylo těžší pozorovat co se děje už v tak složitém prostředí. [Coeckelbergh, 2020]

### 1.3.2 Konekcionismus

Konekcionismus je přístup k vývoji AI, který se vyvinul ze snahy pochopit, jak funguje lidský mozek na neuronové úrovni. Zejména jak je schopný se učit nové věci a pamatovat si je. Vývoj tímto směrem započali neurofyziolog Warren McCulloch z Illinoiské univerzity a Walter Pitts z Chicagské univerzity v roce 1943, když publikovali pojednání o neuronových sítích a automatech. Představili myšlenku, že každý neuron v mozku je jednoduchý digitální procesor a mozek jako celek tak je určitou formou stroje. [Nilsson, 2009: 34-36]

První implementaci této myšlenky vytvořili v 1954 Belmont Farley a Wesley Clark z MIT a to umělou neuronovou sítí. Díky omezené paměti mohla tato síť disponovat maximálně 128 neuronů. Podařilo se jim zprovoznit algoritmus rozpoznávající jednoduché vzory a také zjistili, že pokud náhodně 10 % neuronů zničí, na funkčnost sítě to nemá žádný vliv (mozek také dokáže tolerovat jisté poškození způsobené nehodou či nemocí). Dalším důležitým okamžikem bylo vytvoření jednoho z prvních umělých neuronů (tzv. perceptronu) Frankem Rosenblattem v roce 1958. Přelomovost byla dána tím, že tento neuron bylo možné učit klasifikovat. [Nilsson, 2009: 92-97]

Výzkum perceptronů na nějaký čas skončil s publikací knihy Marvinu Minskyho *Perceptrons* (1969), po které přestali vládní agentury výzkum financovat. Vedla k tomu domněnka, že se perceptron nedokáže naučit logický operátor XOR (buď A, nebo B, ale ne obojí), kterou vyvrátila až publikace *Parallel Distributing Processing* in v polovině 80. let



20 století. Toto období se nazývá neural net winter (zima neuronových sítí) a po jeho konci se začal výzkum opět financovat potřebnými zdroji a z konekcionismu se stal jeden z hlavních směrů při vývoji umělé inteligence. Vedl až k hlubokému učení využívaném v moderních programech [Frankish, Ramsey, 2014: 19].

## 1.4 Testování AI

### 1.4.1 Turingův test

Alan Turing definoval tzv. “Turingův test” v článku *Computing Machinery and Intelligence* v roce 1950. Test se zabývá odlišením strojové inteligence od té lidské. Testem je, zda člověk dokáže rozlišit chování, které každá z nich produkuje. Aby stroj byl inteligentní, a prošel tedy testem, musí vykazovat následující schopnosti:

- Zpracování přirozeného jazyka (schopnost úspěšně komunikovat).
- Reprezentaci znalostí (schopnost ukládat to, co ví nebo slyší).
- Automatické uvažování (schopnost používat uložené informace k zodpovídání otázek a vyvozování nových závěrů).
- Strojové učení (schopnost přizpůsobovat se novým okolnostem a odhalovat vzorce).

Jediným důkazem, z něhož mohou tazatelé vycházet, jsou vstupy a výstupy (otázky položené tazatelem a odpovědi, které tazatel později obdrží). Vyslychanou entitu lze tedy chápat jako černou skříňku v tom smyslu, že neznáme její vnitřní strukturu a znát ji nesmíme. [Wooldridge, 2021: 25] Nejčastěji je test kritizován právě kvůli tomu, že spíše testuje lidskou důvěřivost než skutečnou AI. Je v podstatě o oklamání. [Tegmark, 2017: 119]0

Pokročilé datové modely, například GPT-3, se stále přibližují k úplnému splnění Turingova testu. [Piper, 2020] Stroj však stále neví, o čem vlastně píše a v jakém kontextu. [Marcus, Davis: 2020]

### 1.4.2 Argument čínského pokoje

Jedním z nejznámějších kritiků Turingova testu je John Searle. Ten si položil otázku, zda by počítač vůbec mohl chápat kognitivní stavy nebo smysl toho, co dělá. Ve svém článku představil argument “čínského pokoje” - Searl, jenž neumí čínsky sedí v pokoji plném knih a návodů, jak odpovídat na otázky v čínštině. Na základě těchto prostředků je schopný odpovídat na vzkazy, které mu dává osoba mimo pokoj. Je tedy schopný produkovat korektní “výstupy” na “vstupy”, aniž by jim rozuměl. Test tak splnil.

Searle tvrdí, že není přítomno žádné porozumění čínštině. Člověk v místnosti nerozumí čínštině a místnost sama jistě čínštině nerozumí. Jeho lidská inteligence je využita pouze do té míry, že plní zadané pokyny. Dělá stejné věci jako počítač, jednoduše postupuje podle seznamu instrukcí. Podle Searla tedy na základě stejného argumentu počítač, který projde Turingovým testem, nebude vykazovat porozumění. Pokud je Searlův argument správný, znamená to, že silná umělá inteligence nemůže vzniknout následováním receptu. Silné umělé inteligence tedy nelze dosáhnout pomocí běžných počítačů. [Wooldridge, 2021: 166]

### 1.4.3 Winograd Schema Challenge (WSG)

Test strojové inteligence pojmenovaný po profesoru informatiky na Stanfordově univerzitě Terryem Winogradovi, vznikl jako vylepšení Turingova testu. Jedná se o test s výběrem odpovědí, který používá otázky s velmi specifickou strukturou, tzv. winogradských schémat. Zaměřují se přímo na rozumové chápání, které počítačovým systémům v době jeho vzniku často chybělo. [Tegmark, 2017: 119]

Typicky test pokládá otázku na koho se v následujících větách vztahuje skryté zájmeno "oni":

1. Městští radní odmítli demonstrantům povolení, protože (oni) se báli násilí.
2. Městští radní odmítli demonstrantům povolení, protože (oni) obhajovali násilí.

Otázky a první pohled vypadají jako úloha zpracování přirozeného jazyka. Stroj musí určit jaké slovo předchází zájmenu. Dle tvůrce testu však jde o využití znalostí a rozumového uvažování. [Levesque, 2014]

Moderní modely AI již s WSG nemají ve většině případů problémy a vykazují úspěšnost nad 90 % bez specifických úprav. Pokud použijeme překladáč DeepL na anglickou verzi příkladu výše [Tegmark, 2017: 119], získáme fakticky správné věty v českém jazyce.

- *"The city councilmen refused the demonstrators a permit because they feared violence."*
  - "Radní odmítli demonstrantům vydat povolení, protože se obávali násilí."
- *"The city councilmen refused the demonstrators a permit because they advocated violence."*
  - "Městští radní odmítli demonstrantům vydat povolení, protože obhajovali násilí."

#### 1.4.4 Lovelace test

Lovelace test, pojmenovaný po matematicce Adě Lovelace z 19. století, vychází z myšlenky, že pokud se chceme u umělé inteligence zabývat schopnostmi podobnými lidským, nesmíme zapomínat na schopnost vytváření nových, originální věcí.

Test tedy používá kreativitu jako zástupce inteligence. Byl navržen v roce 2001 a jeho nosnou myšlenkou byl agent, který by něco vytvářel - například příběh nebo báseň. Aby agent tímto testem prošel, musel by vytvořit kreativní výstup takovým způsobem, aby tvůrce agenta nedokázal vysvětlit, jak tento výstup vytvořil. Pokud je vytvořený výstup něco úplně nového a překvapivého, musí mít stroj něco, co připomíná vědomí. [Pearson, 2014]

Hlavním problémem testu tvořila část o neschopnosti vysvětlení způsobu, jak agent výstup vytvořil. Je velice nepravděpodobné, že by programátor nedokázal přijít na to, jak jeho umělá inteligence něco vytvořila. Na základě původního Lovelace testu navrhl Mark Riedl v roce 2014 jinou verzi – Lovelace 2.0 Test. Test se liší tím, že hodnotitel vědomě komunikuje s umělou inteligencí. Zadává jí úlohy o dvou složkách – kreativní artefakt a kritérium. Například “Řekni mi příběh o kočce, která šla na nákup” nebo “Nakresli mi obrázek muže, který drží tučňáka.”

Pokud je hodnotitel s výsledkem spokojen, zadá znovu o něco těžší úlohu. Takto pokračuje až do doby, než vyhodnotí umělou inteligenci jako úspěšnou (prokáže dostatečnou inteligenci) nebo naopak neúspěšnou. Iterativní vyhodnocování umožňuje porovnávat výkony jednotlivých systému pomocí skóre úspěšné/neúspěšné. [Riedl, 2014]

V letošním roce (2022) byl představen nový model na generování obrázků DALL-E 2. Riedl se při té příležitosti vyjádřil, že se mění i představy o tom, co rozumíme pod pojmy “vytvořit” a “pochopit”: “Tyto pojmy jsou špatně definované a jsou předmětem diskusí. Například včela chápe význam žluté barvy, protože na základě této informace jedná. Pokud definujeme chápání jako lidské chápání, pak jsou systémy umělé inteligence velmi daleko.” Dále dodal [Riedel, 2014]: “Ale také bych tvrdil, že tyto systémy pro generování umění mají určité základní porozumění, které se překrývá s lidským porozuměním. Dokážou případně nasadit „sukýnku na ředkvičku“<sup>5</sup> na stejné místo, kam by ji nasadil člověk.” Stejně jako včela, i DALL-E 2 pracuje s informacemi a vytváří obrazy, které splňují lidská očekávání a nutí nás přemýšlet, co si pod těmito pojmy vlastně představujeme. [Heaven, 2022]

## 1.5 Historie vývoje umělé inteligence

Popsat celou historii umělé inteligence (AI) není jednoduchý úkol a pravděpodobně by na něj mohlo vzniknout několik akademických prací. Už jen vymezení samotného oboru není jednoduché – představy o umělé inteligenci se datují do vzdálené minulosti, ať už jde o Pražského Golema nebo o kovové tvory z řecké mytologie. Pro účely této práce jsem se rozhodla vývoj umělé inteligence vylíčit hlavně pomocí událostí, které vedly k transformaci v počátcích filozofického oboru do ryze technického, jak ho známe teď.

Vzhledem k zaměření práce spíš na současnost a budoucnost jsem se rozhodla začít příběhem Alana Turinga, který kromě velkého vlivu na konec druhé světové války také založil počítačovou vědu.

### 1.5.1 Nové pojetí stroje a vznik oboru

V počátcích moderních dějin byl stroj vnímán jako označení pro něco, co operuje vždy stejně, s daným výstupem a podle předem daných pravidel – to můžeme ilustrovat na příkladu rčení “šlape jako hodinky” používaného, když něco spolehlivě funguje stále stejným způsobem delší dobu. Toto platilo ještě do nedávna i pro počítače – na začátku 50. let 20. století stále byly tímto přesně programovatelným a předvídatelným zařízením. V ten moment se začaly počítače vyvíjet v úplně nový druh stroje, začalo se mluvit o jejich inteligenci. Takový stroj je ve svém základu diametrálně jiný, než co lidstvo znalo a rozlišovalo před tím. [Nilsson, 2009: 81-85]

Pravděpodobně žádný z pokroků v této oblasti by nebyl možný bez práce a myšlenek britského matematika Alana Turinga. Turing je po právu nazýván za “otce informatiky” a jeho klasická esej *Computing Machinery and Intelligence* zkoumá možnosti počítačů vytvořených pro simulaci jevů dnes spojovaných s AI včetně toho, jak by se stroje mohli učit a jak umělou inteligenci testovat. Bohužel, ve své době tyto myšlenky byly pouze teorie - Turing neměl výpočetní prostředky k tomu, aby je mohl převést do praxe a obor zabývající se inteligentními stroji tak formálně vznikl až 2 roky po jeho smrti. O to více pozoruhodné však je, že i tak dokázal inspirovat jeden z nejvíce technologicky pokročilých oborů dnešní doby. [Coeckelbergh, 2020: 40-43]

Název umělá inteligence a stejnojmenný obor oficiálně vznikl na letním semináři, který uspořádal John McCarthy v roce 1956 v rámci *Dartmouth Summer Research Project on Artificial Intelligence*. Seminář se konal v Hanoveru, New Hampshire a jeho hlavním cílem bylo prozkoumat způsoby, kterými by bylo možné docílit aspektů inteligence u strojů. Tuto myšlenku můžeme považovat za stěžejní v celém oboru. První použití termínu se

připisuje samotnému Johnu McCarthymu, který ho použil právě v návrhu na uspořádání tohoto semináře, vytvořeném společně s Marvinem Minskym, Nathanielem Rochesterem a Claudem Shannonem. [Nilsson, 2009: 77-80]

Diskusí o strojích, které “mohou myslet”, si všimli i v tehdejší Sovětském svazu. Novináři psali o skrytých snahách kapitalistů nahradit stávkující dělníky roboty a stejně tak piloty, kteří nechtěli bombardovat civilisty, “lhostejnými kovovými monstry”. Sovětští filosofové napadali celou vizi “strojů co umí přemýšlet” jako “idealistickou”, kvůli faktickému odtržení od lidského mozku a “mechanickou” kvůli realizování skrz počítačové operace. Kritici obecně radili veškeré “kybernetické” využití počítačů jako “idealistickou pseudovědu” a celkově shazovali její smysl, zároveň vývoj v té oblasti považovali za vojenskou provokaci. Tento přístup a mediální rozruch zapříčinil zmrazení vývoje této oblasti v Sovětském svazu z ideologických důvodů. [Gerovitch, 2011: 173–194]

Ačkoliv na semináři v Dartmouth College oficiálně vznikla výzkumná komunita a identita oboru, některé technické myšlenky charakteristické pro umělou inteligenci můžeme nalézt již před jeho konáním. Thomas Bayer již v 18. století zkoumal pravděpodobnost jevů, George Boole v 19. století dokázal systematické využití Aristotelovy logiky stejným způsobem jako lze řešit soustavu rovnic. Na přelomu 20. století vedl pokrok v experimentálních vědách k rozvoji statistických metod, umožňujících hledání souvislostí ve větším množství dat. V 50. letech 20. století pak byly zkonstruovány první elektronické počítače a primitivní autonomní roboti. [Nilsson, 2009]

Mezi 50. a 70. lety 20. století bylo učiněno množství objevů v několika zásadních oblastech umělé inteligence. Herbert Simon a Allen Newell se zabývali heuristickým vyhledáváním a jejich program *Logic Theorist* je považován za vůbec první disponující umělou inteligencí. Program hledal logické důkazy na teorémy z knihy *Principia Mathematica* - byl jich schopen vyřešit 38 z 52. U jednoho teorému dokonce přišel s důkazem, který byl elegantnější než původní řešení - tímto vyvrátil představu, že stroje dokážou "myslet pouze v číslech" a dokázal schopnost dedukce a vymýšlení logických důkazů. Později pak vytvořili *General Problem Solver* který kromě řešení matematických důkazů uměl i hrát šachy. Oliver Selfridge se svým týmem pracovali na rozpoznávání znaků a položili tak základ oblasti počítačového vidění, ve které začali vznikat čím dál více sofistikovanější programy, například se schopností rozpoznat lidské obličej. [Nilsson, 2009]

Dalším významným mezníkem byl první chatbot ELIZA, vytvořeny v letech 1964 až 1966 Josephem Weizenbaumem na MIT. Program byl také první, který prošel Turingovým

testem ačkoliv autor takovou “zkoušku” nikdy nezamýšlel a pouze to rozpoutalo diskuzi o vhodnosti a smyslu celého testu. ELIZA měla roli psychiatra, který hovoří s pacientem a klade mu otevřené otázky, aby ho přiměl k zamyšlení nad sebou samým. ELIZA pracovala tak, že hledala v otázkách klíčová slova (jako je rodina, přátelé, osamělý, smutný apod.) a poté pomocí připravených skriptů spojených s příslušným slovem pokládala doplňující otázky. Nelze tedy v žádném případě mluvit schopnosti chápání textu a nějaké pokročilé inteligenci. [Coeckelbergh, 2020: 25-26]

### 1.5.2 První AI Winter

Po sérii úspěchů byla na začátku 70. let 20. století vědecká komunita frustrovaná neúspěchy a pomalým vývojem v oblasti základních problémů AI a fantastických prohlášení některých vědců. Nespokojenost kulminovala v polovině 70. let a jedním z nejhlasitějších kritiků byl americký Filozof Hubert Dreyfus, kterého pověřila společnost RAND Corporation, aby sepsal současný stav vývoje umělé inteligence. Jako název pro jeho publikaci si zvolil “*Alchemy and AI*” a dal tím tak jasně najevo svůj názor na celou oblast – název vycházel z toho, že konečné cíle AI jsou stejně nedosažitelné tak, jako ty v alchymii. Ačkoliv se v dnešní době můžeme dívat na práci negativně, nelze Dreyfusovi upřít několik správných připomínek, zejména z oblasti nadsazených tvrzení a nerealistických předpovědí průkopníků. Dnes již víme že byli opravdu neuvěřitelně nerealistické. [Coeckelbergh, 2020: 52]

Zmínit můžeme následující [Simon, 2017]:

- 1958, Herbert Simon a Allen Newell: "Do deseti let bude digitální počítač mistrem světa v šachu."
- 1965, Herbert Simon: "Stroje budou do 20 let schopny vykonávat jakoukoli práci, kterou může vykonávat člověk."
- 1967 Marvin Minsky: "Během jedné generace bude problém vytvoření 'umělé inteligence' v podstatě vyřešen."
- 1970 Marvin Minsky: "Za 3 až 8 let budeme mít stroj s obecnou inteligencí průměrného člověka."

Na druhé straně oceánu byl pověřen obdobným úkolem profesor z univerzity v Cambridge James Lighthill. Jeho zpráva nazvaná “*Artificial Intelligence: A General Survey*” poněkud nezvykle rozdělovala výzkum AI na tři kategorie – pokročilá automatizace, počítačové studie a studium centrálního nervového systému. Nejvíce kritický byl ke

kategorii počítačových studií, která obsahovala obecný výzkum AI, robotiku a zpracování přirozeného jazyka. Ve své studii napsal “V žádné z oblastí nepřinesly dosavadní objevy takový dopad, jaký byl tehdy (kolem roku 1960) slibován.”<sup>6</sup>. Zprávu uzavřel tvrzením, že existující techniky AI, které fungovaly v laboratorních podmínkách, nebude možné škálovat na reálné problémy kvůli kombinatorické explozi<sup>7</sup>. Tato zpráva vedla k drastickým škrtům financování výzkumu po celém Spojeném Království. [Nilsson, 2010: 261-264]

Ve Spojených státech byla v té době hlavním sponzorem výzkumu vojenská organizace DARPA (a její předchůdce ARPA), zde však zavládla také frustrace z nesplněných slibů a vlivem války ve Vietnamu bylo financování značně omezeno. Období od 70. do 80 let se později začalo označovat jako *AI Winter*. Přesnější by bylo označení *první AI winter*, jelikož další měli následovat. Také vznikl stereotyp o AI jakožto o oboru, který vytváří přehnaně optimistické a neopodstatněné předovědi a sliby, které není schopen naplnit. V komunitě vědců umělá inteligence získala pověst podobnou homeopatické medicíně a zdálo se, že jako seriózní vědecká disciplína je na pokraji úpadku. [Wooldridge, 2021: 53]

### 1.5.3 Vzestup expertních systémů

Ačkoliv mezi akademiky stále kolovala Lighthillova zpráva a mnozí ho považovali AI za pseudoobor, nová generace vědců z konce 70. a začátku 80. let 20. století začalo prosazovat nový přístup. Ten měl řešit nedostatky těch předchozích a předejít situacím, které vrhaly celý obor do špatného světla. Hlavní překážkou bylo zaměření AI na příliš obecné věci jako bylo hledání řešení obecných problémů nebo matematické úlohy. Na řešení těchto oblastí chyběl jeden ze základních prvků inteligence - znalosti. Zastánci tohoto nového přístupu věřili, že hlavním prostředkem pokroku je zachycení lidského vědění a předáního do počítačového programu. Vznikla tak nová odnož - expertní systémy. [Wooldridge, 2021]

Cílem expertního systému je za pomoci vložených znalostí a pravidel řešit úzce specifikované problémy. Úspěšné použití expertních systému dokázalo, že AI dokáže v určitých oblastech předčít člověka. A co bylo důležitější, dokázali komerční využitelnost AI ke generování zisku. Tyto systémy se nesnažily být obecnou inteligencí, tak jak si ji představoval Turing. Typickým příkladem byly oblasti, ve kterých existovaly dlouho sbírané znalosti, například v medicíně. Prvním takovým úspěšným systémem byl MYCIN, který

---

<sup>6</sup> “In no part of the field have the discoveries made so far produced the major impact that was then (around 1960) promised.”

<sup>7</sup> neformální označení jevu v matematice, kdy složitost daného problému silně vzrůstá spolu s tím, jak se vzhledem k rostoucímu vstupu velice rychle rozšiřuje kombinatorické jádro problému, typicky počet kombinací, které by mohly být řešením

dokázal převahu stroje nad člověkem v určité oblasti. MYCIN, který vznikl na univerzitě ve Stanfordu, měl být asistentem ošetřujícího lékaře zajišťující poradenství ohledně léčby. Hlavní příčinou jeho úspěchu bylo to, že lidé kteří ho vyvíjeli skutečně byli experty v dané oblasti (část podobných systému v budoucnu na nedostatečné odbornosti jejich tvůrců doplatila a nebyla úspěšná). [Nilsson, 2017:292-295]

Znalosti byli zaneseny pomocí pravidel *jestliže – pak*<sup>8</sup> a během 5 letého “zanášení” jich v systému existovaly stovky. Úspěch a status “ikonického systému” byl dán zejména tím, že MYCIN splňoval všechny základní předpoklady funkčního expertního systému:

- Fungování připomínalo konzultaci s lidským expertem – sekvence otázek na uživatele a následná diagnóza se stala základem všech expertních systémů.
- MYCIN byl schopný zdůvodnit své rozhodnutí – transparentnost toho, jak systém došel k rozhodnutí se stalo zásadním. V případě systému MYCIN mohli být následky dalekosáhle až smrtelné. Ti, kteří se systémem řídili museli mít jistotu a pro tu je zdůvodnění nezbytné. Lidé obecně nevěří systémům, které dávají odpovědi ve formě tzv. “black boxu” a nezdůvodní je.
- Systém byl schopen zodpovědět otázky ohledně toho, jak dospěl k rozhodnutí – toho bylo docíleno vytvořením řetězce úvah, které vedly k závěru – pravidla, která byla spuštěna, a informace, které byly získány a které vedly ke spuštění těchto pravidel. Ačkoli taková vysvětlení nejsou ideální, lze je snadno vytvořit a poskytují alespoň nějaký mechanismus, jehož prostřednictvím lze pochopit závěry systému.
- MYCIN se dokázal vypořádat s nejistotou, tedy situacemi, ve kterých nešlo s jistotou říci, že poskytované informace jsou pravdivé. Zpracování nejistoty se ukázalo jako základní požadavkem na expertní systém a obecněji pro všechny systémy umělé inteligence. V systémech, jako je MYCIN, lze jen zřídka na základě konkrétního důkazu učinit jednoznačný závěr. Například, když se pacientovi provede krevní test, který je pozitivní, pak to poskytuje určitou informaci, kterou systém může zohlednit, ale vždy bude existovat možnost chybného testu (falešně pozitivního nebo falešně negativního). Případně pokud se u pacienta projeví nějaký příznak, pak to může svědčit o určitém onemocnění, ale nestačí to k tomu, aby bylo možné s jistotou uzavřít, že pacient tímto onemocněním trpí. Aby byl expertní systém schopen činit správné rozhodnutí, musel by dokázat takové důkazy zásadním způsobem zohlednit.

Testy provedené v roce 1979 na deseti skutečných případech ukázaly že výkonnost systému MYCIN při diagnostice krevních chorob je srovnatelná s lidskými odborníky. V případě praktických lékařů byla dokonce vyšší. Jednalo se o jeden z prvních případů, kdy systém umělé inteligence prokázal schopnosti na úrovni lidských expertů nebo lepší v úkolu ze skutečného světa. [Wooldridge, 2021:57]

---

<sup>8</sup> V originále IF – ELSE – typ větvení algoritmu



#### 1.5.4 Ekonomická bublina a druhá AI Winter

Po počátečních úspěších expertních systémů vznikla celá řada komerčních firem, zabývajících se vývojem AI. Mnozí věřili, že vkládání expertních znalostí a používání pravidel “jestliže – pak” je ta nejlepší cesta k vývoji umělé inteligence. Expertní systémy se prosadili v celé řadě odvětví, velmi časté byli ve finančním plánování, lékařské diagnostice, geologickém průzkumu a k návrhu mikroelektronických obvodů. [Nilsson, 2009: 285-303] Časopis Business Week se k tomuto boomu se připojil a v roce 1984 zveřejnil titulky "AI: Je to tu?". [Schuchmann, 2019] Spousta firem také předpovídala neskutečné pokroky, když jich pak nedokázali dosáhnout a dodat slibované produkty, veřejné mínění ohledně AI se opět začalo zhoršovat. Vědci z oboru se začali obávat přehnaného optimismu a na konferenci asociace AAI (Association For The Advancement Of Artificial) v roce 1984 během panelové diskuze s názvem “Doba temna umělé inteligence - Můžeme se jim vyhnout, nebo je musíme přežít?” se snažili najít způsob, jak předejít další AI Winter. Panoval zde obava, zda přehnané a opět nesplněné sliby nepovedou k přerušování financování výzkumu. Obavy se bohužel vyplnily. [Newquist, 1994]

V následujících letech se objevovalo stále více kritiky schopností systémů AI, zejména těch expertních – revolučního vynálezu posledního “léta” umělé inteligence, který obnovil zájem o obor. John McCarthy v roce 1984 zveřejnil kritiku expertních systémů kvůli absenci “selského rozumu” a jejich vlastní nevědomosti o omezeních, která mají. Problém popsal, na již zmíněném systému MYCIN a nastínil situaci, kdy má pacient cholera. Systém předepsal pouze antibiotika *tetracycline* na dva týdny a nic více. Léčba by pravděpodobně zabrala, avšak pacient by byl již tou dobou po smrti. Lidský lékař by jistě věděl, že průjem způsobený cholerou je nutné léčit jiným způsobem. Další z problémů, který vyvstala byla drahá a složitá údržba existujících systémů. Pokud navíc počet pravidel překročil určitou hranici, stávala se z nich “černá skříňka”. Některé programy měly tak složitá pravidla, že je nebylo téměř možné zadat ručně. To platilo zejména pro systémy rozpoznání řeči a pro strojové vidění. Jednoduše měly příliš mnoho mezních případů. Situaci paradoxně nepomohl ani rozvoj osobních počítačů, který způsobil krach trhu se specializovaným hardwarem v roce 1987. Stolní počítače společností Apple a IBM se výkonově dotáhly na mnohem dražší specializované stroje pro AI a nebyl tedy důvod je nadále kupovat. Celé jedno odvětví, v hodnotě půl miliardy dolarů, bylo přes noc víceméně zničeno. [Newquist, 1994]

Nové vedení americké organizace DARPA razantně snížilo financování výzkumu na konci 80. let 20. století s vysvětlením, že AI není ta “další velká věc” a přesunulo finance na projekty přinášející okamžité výsledky. Celá druhá zima AI by se tedy dala popsat jevem

ekonomické bubliny – významné odpoutání reálné vnitřní hodnoty aktiva od ceny, za kterou se aktivum obchoduje. V případě AI šlo o nadhodnocení celého odvětví příliš optimistickými odhady budoucnosti. Ke konci roku 1993 tak ukončilo činnost, zkrachovalo nebo bylo převzato více než 300 firem z oblasti umělé inteligencí. Fakticky tak skončila první vlna komerční AI. [Newquist, 1994: 400]

### 1.5.5 Devadesátá léta, neuronové sítě a Gary Kasparov

Je téměř až ironií, že znatelný pokrok v oblasti umělé inteligence v 90. letech 20. století se odehrál bez financování vládními agenturami a bez většího zájmu veřejnosti. Části těchto pokroků bylo dosaženo výkonnější technikou, část cílením na užší a lépe definované problémy. Pošramocená pověst a malá shoda ohledně příčin neúspěchu uvnitř celé oblasti vedly k rozdělení na menší, více specializované obory (kognitivní systémy, inteligentní systémy, reprezentace znalostí apod.). Někteří dokonce odstranili jakékoliv vazby na umělou inteligenci jako takovou. [Wooldridge, 2021]

Optimismus zavládl zejména díky novým technikám a přístupům, které byly alternativou k dosud nejčastějšímu symbolickému přístupu používaném v expertních systémech. Staronové techniky jako neuronové sítě nabízely řešení klasických problémů expertních systémů, například jejich *křehkost*. Nesmyslný výstup, pokud programátor udělal i sebemenší chybu v logice. Jejich větší "organičnost" zajišťovala větší toleranci k chybám, nefungovala-li část systém produkoval "jen" horší výstupy místo toho aby byl kompletně nefunkční. [Torres-Huitzil, Girau, 2017]

K zásadní události došlo v roce 1997, kdy novinové titulky zaplnila zpráva o výhře počítače *Deep Blue* společnosti IBM nad úřadujícím mistrem světa v šachu Garry Kasparovem. Poprvé Kasparov s počítačem hrál již v lednu 1996, kdy sice počítač vyhrál první hru, ale celé utkání následně prohrál. V květnu 1997 však hardwarově silnější verze, neoficiálně označována jako *Deeper Blue*, mistra svět porazila. Po porážce Kasparov podle *New York Times* prohlásil: "Neměl jsem vůbec náladu hrát." V článku se dále píše, že "po páté sobotní partii byl tak rozladěný, že měl pocit, že zápas už skončil. Na otázku proč, odpověděl: "Jsem lidská bytost. Když vidím něco, co je daleko za hranicí mého chápání, mám strach". [Webber, 1997] Verze počítače, která Kasparova porazila měla 256 speciálních "šachových procesorů" a byla schopná vyhodnotit 200 milionů šachových pozic za sekundu – Kasparov pouze 3. Kasparov se spoléhal na své znalosti a intuici, zatímco stroj, který nic takového neměl, pouze na hrubou sílu. [Nilsson, 2009: 591-595]

Deep Blue nevyužíval žádné pokročilé technologie té doby (strojové učení, neuronové sítě apod.), byl pouze jakýmsi vrcholným představitelem expertního systému a symbolické AI. Systém měl “naučeno” velké množství šachových partií (včetně všech Garyho Kasparova) a dokázal tak vyhodnotit aktuální tah a předpovědět ten další, zejména díky na tu dobu neskutečně velkému výpočetnímu výkonu. Deep Blue neuměl “myslet”, nemohl vymýšlet nové strategie, nebo se učit ze svých partií. Z hlediska vývoje AI tedy o nic přelomového nešlo, z hlediska historie a milníků ano. Také se objevily spekulace o tom, zda společnost IBM zápas nezmanipulovala a nechtěla jen získat publicitu a zvýšit tak hodnotu svých akcií, což se opravdu stalo, a to dokonce o 10 %. [Newborn, Lierson: 2002]

Během tohoto období bylo přijato nové paradigma nazývané “inteligentní agenti”. Agent je systémem umělé inteligence, který je samostatnou entitou nacházející se v určitém prostředí a vykonává určitý úkol, který vykonává “jménem” uživatele. Zaměření se na kompletní systém, a nikoliv jen na jednotlivé složky inteligence (učení se, uvažování apod.) byl významný krok k moderním systémům, které známe dnes. Po období debat se vědecká komunita shodla na třech důležitých schopnostech inteligentního agenta [Wooldridge, 2021: 75-79]:

1. Agent musí být reaktivní – musí se přizpůsobit svému prostředí, pokud v něm nastanou změny.
2. Agent musí být proaktivní – musí být schopen systematicky pracovat na splnění úkolu jménem uživatele.
3. Agent musí být sociální – musí umět spolupracovat s jinými agenty, když si to situace žádá.

Symbolická AI zdůrazňovala důležitost proaktivního chování (plánování a řešení problémů), zatímco behaviorální AI zdůrazňovala důležitost reaktivního chování (vtělení se do prostředí a přizpůsobení se mu) - agenti tvořili jakýsi průsečík těchto dvou přístupů a přinášeli revoluční myšlenku jejich kooperace. Využití principu agentů ve světě softwaru začalo být velmi zajímavé ve spojení s rozvojem internetu, který také probíhal v 90. letech 20. století. Jako průkopník by se dal označit vědec belgického původu Pattie Maes z MIT, který publikoval článek popisující experimentální software v jeho laboratoři využívající princip agentů. Jeho emailový asistent se učil chování uživatele při zpracování emailů (které přednostně čte, které rovnou maže, které organizuje do složek apod.) a když si byl dostatečně “jistý”, začal tyto činnosti provádět sám. Tyto programy byly předobrazem nejen osobních asistentů, ale celého směru, jakým se obor nadále ubíral. [Wooldridge, 2021]

### 1.5.6 Autonomní vozidla a hluboké učení

Alan Turing předpovídal, že v roce 2000 v rozhovoru na určité téma bude stroj nerozeznatelný od člověka. V témže roce však neexistoval žádný počítač, který by dokázal po 5 minutách konverzace přesvědčit. Jeho proroctví se naplnilo o několik desítek let později, částečně a v pravém slova smyslu žádný počítač ještě “Turingovým testem” neprošel. Každý rok se koná soutěž o “Loebnerova cenu” která tento test bere doslovně. Porota vede rozhovory přes chatovací program a snaží se odhadnout, zda na druhém konci je člověk nebo umělá inteligence. Soutěž je často kritizována, jelikož programy používají triky jako umělé vytváření pravopisných chyb nebo vtípkování místo snahy o skutečnou umělou inteligenci. Velmi často soutěží jen upravené verze programu ELIZA ze 70. let 20. století. Žádný z programů zatím cenu nezískal, i když některé byly schopny oklamat porotu i při delší konverzaci. [Wooldridge, 2021: 25-27]

John McCarthy snil na konci 50. let o tom, že by ho jednou mohlo na letiště dovézt autonomní vozidlo. V roce 1987 se tomu přiblížil Ernst Dickmanns, který předvedl na prázdné německé dálnici samořídící dodávku Mercedes v rámci projektu Prometheus. Dodávka poměrně obstojně zvládla běžnou jízdu. Vrcholem projektu Prometheus byla pak v roce 1995 cesta z německého Mnichova do dánské Kodaně a zpátky. Bez zásahu řidiče dodávka ujela až 158 km a dosáhla rychlosti přes 175 km/h. [Nilsson, 2009: 604]

Série pokroků přiměla americkou agenturu DARPA uspořádat soutěž autonomních vozidel s názvem Grand Challenge. Cílem bylo projet 150 mil (necelých 242 km) americkou krajinou a hlavní výhra byl milion amerických dolarů. Přihlásilo se 106 týmů a organizace vybrala 15 finalistů – žádný z nich však závod nedokončil. Některá vozidla nevyjela ani ze startu a nejúspěšnější vozidlo Sandstorm z univerzity Carnegie Mellon urazilo 12 kilometrů, než vyjelo z trati a uvízlo na náspu. [Nilsson, 2009: 604-611]

Ačkoliv se závod mohl jevit jako fiasko, DARPA ohlásila další kolo soutěže na příští rok a zdvojnásobila výhru na 2 miliony amerických dolarů. Na druhý ročník se přihlásilo 195 soutěžících a bylo vybráno i více finalistů - 23. Cílem bylo urazit 132 mil (necelých 213 km) v nevadské poušti. Vítězem se stalo vozidlo Stanley ze Stanfordské univerzity. Stanley byl upravený Volkswagen Tourage se 7 počítači zpracovávající data ze senzorů GPS, laserových dálkoměrů, radaru a kamer. Trať urazil za necelých 7 hodin a v průměru jel rychlostí 20 milů za hodinu (32 km za hodinu). Úspěch závodu byl velkým milníkem a lze ho přirovnat k prvním letu letadla Kitty Hawk bratří Wrightů. Ačkoliv let trval jen 12 sekund dokázal, že je možné, aby se stroj těžší než vzduch vznesl a letěl [Wooldridge, 2021: 117] Grand Challenge udělal z autonomních vozidel realitu. [ibid.]

Na poli neuronových sítí nebyla na počátku milénia situace pro výzkumníky příliš příznivá díky poslední AI winter. Výzkumníci Geoffrey Hinton, Yann LeCun and Yoshua Bengio v jejich potenciál stále věřili a nebyli spokojeni se situací, kdy jejich vlastní výzkum ale i ten studentů není publikován jen kvůli tématu neuronových sítí. Výzkum v oblasti strojového učení a umělé inteligence byl stále aktivní a s neuronovými sítěmi stále pracovali i další výzkumníci. Počty citací z té doby ale jasně ukazují, že zájem zcela nezmizel, nicméně razantně opadl. Silného spojence našli v Kanadské vládě, kam se Hinton, někdy přezdívaný kmotr výzkumu neuronových sítí, přestěhoval v roce 1987 právě kvůli financování jeho výzkumu z programu CIFAR (Canadian Institute for Advanced Research). Ačkoliv bylo tento program ukončen v polovině 90. let, Hinton dokázal získat financování na fungování menšího týmu a mohl pokračovat ve výzkumu. [Foote, 2022]

Sám Hinton vypráví o zosnování “spiknutí” a přejmenování oboru na “hluboké učení” aby se vyhnuli negativním konotacím spojených s neuronovými sítěmi z minulosti. Hinton, Simon Osindero a Yee-Whye Teh publikovali v roce 2006 článek, který byl považován za průlomový a dostatečně významný na to, aby oživil zájem o neuronové sítě: *A fast learning algorithm for deep belief nets*. Tento článek přinesl revoluční myšlenku propojení několika neuronových sítí a položil tak základy neuronovým sítím tak, jak je známe teď. Základní myšlenky tohoto využití se datují do 90. let. V té době kvůli nedostatku digitálních dat a výkonu počítačů nemohly být realizovány a dále zkoumány. [Wooldridge, 2021: 95-96]

Většina průlomů v posledních letech byla možná právě díky těmto vědcům, kteří stále věřili myšlence a potenciálu emulování lidského mozku. Michael Wooldridge ve své knize vzpomíná [Wooldridge. 2021: 86]: “V červenci 2000 jsem na konferenci v Bostonu sledoval prezentaci jedné z mladých hvězd nové umělé inteligence. Seděl jsem vedle zkušeného veterána v oblasti umělé inteligence – někoho, kdo se jí věnoval už od zlatého věku, současníka McCarthyho a Minského. On byl opovrhlivý. "Tohle se dnes považuje za umělou inteligenci?" zeptal se. "Kam se podělo to kouzlo?" Chápal jsem, z čeho vychází: kariéra v oblasti umělé inteligence dnes vyžaduje vzdělání nikoli ve filozofii, kognitivní vědě nebo logice, ale v pravděpodobnosti, statistice a ekonomii. To už nevypadá tak poeticky, že?”<sup>9</sup>. Tato vzpomínka krásně ilustruje, jak se v novém tisíciletí celý obor transformoval.

---

<sup>9</sup> In July 2000, I was at a conference in Boston, watching a presentation by one of the bright young stars of the new AI. I was sitting next to a seasoned AI veteran—someone who had been in AI since the golden age, a contemporary of McCarthy and Minsky. He was contemptuous. “Is this what passes for AI nowadays?” he asked. “Where did the magic go?” And I could see where he was coming from: a career in AI now demanded a background not in philosophy or cognitive science or logic but in probability, statistics, and economics. It doesn’t seem as, well, poetic, does it?

Autoři původních myšlenek mohli být zklamaní, že po skutečné realizaci to není tak “magické”, jak si představovali a stále se jedná o počítačové programy.

### 1.5.7 Watson, Tesla, AlphaGo a pokroky ve strojovém učení

Zájem veřejnosti o AI v druhé dekádě odstartoval další ze strojů IBM – Watson. V únoru 2011 porazil dva nejúspěšnější šampiony americké televizní soutěže Jeopardy! - Kena Jenningse a Brada Ruttera. Třídenní exhibiční zápas deník The New York Times označil za “ospravedlnění akademického oboru umělé inteligence”<sup>10</sup>. [Markoff, 2011] Schopnost stroje reagovat a porozumět nápovědám (v soutěži nejsou klasické otázky ale slovní rébusy na které se odpovídá otázkou) byla považována za obří krok kupředu v oblasti zpracování přirozeného jazyka a vyhledávání informací. Watson nebyl během soutěže připojen k internetu. Společnost IBM posléze také oznámila, že stroj poskytne lékařům na pomoc při stanovování diagnóz a léčbě pacientů. [Gammon, 2011]

Soutěž Jeopardy! byla pro umělou inteligenci opravdovou výzvou. Nápovědy jsou spíše "podivné, krátké haiku, protkané narážkami, slovními hříčkami, pomrkáváním a chytáky na odvedení pozornosti". [Jennings, 2019] Pokud Watson chyboval, bylo to kvůli “nepochopení” pro člověka jasných nápověd. Mnozí se domnívali, že Watson vyhrál jen díky své operační paměti o kapacitě 15TB<sup>11</sup> a vložených datech z milionů dokumentů, knih, encyklopedií a zpravodajských článků. [IBM, 2012] Nabízí se však mnohem jednodušší aspekt poskytující Watsonovi největší výhodu ze všech. Byl mnohem rychlejší než jeho lidské konkurenti. Během soutěže znali odpovědi pravděpodobně všichni zúčastnění, vyhrával však ten, kdo nejrychleji zmáčkl tlačítko. Vědec z IBM Eric Brown na kritiku ohledně výhody “v mačkání tlačítka” reagoval: “V některých věcech budou počítače lepší než lidé a naopak. Lidé jsou mnohem lepší v porozumění přirozenému jazyku. Počítače jsou lepší v reagování na signály.”. [Gustin, 2011] Úspěchu Watsona využil například i futurista Martin Ford k otevření již klasického tématu problémů pracovního trhu spojených s příchodem AI - “Koneckonců, pokud stroj dokáže porazit člověka v soutěži Jeopardy! budou počítače brzy soutěžit s lidmi o pracovní místa založená na znalostech?”<sup>12</sup>. [Ford, 2011]

V současné době je pravděpodobně pro veřejnost nejdostupnější systém autonomního řízení vozu program *Autopilot* od společnosti Tesla. První verze Tesla Model

---

<sup>10</sup> “a vindication for the academic field of artificial intelligence”

<sup>11</sup> 15000 Gb – většina dnešních laptopů disponuje od 8 - 16gb

<sup>12</sup> “After all, if a machine can beat humans at Jeopardy!, will computers soon be competing with people for knowledge-based jobs?”

S vznikla v roce 2012 a v té době šlo o technicky nejvyspělejší automobil na běžném trhu. Od září 2014 všechny vozy Tesla Model S měly ve výbavě kamery, radar a akustické snímače dosahu. Účel této výbavy se však stal jasný až o rok později v říjnu 2015. Tesla zveřejnila aktualizaci operačního systému vozu, který obsahoval novou funkci *Autopilot*. Autonomní jízdu však měl ještě poměrně omezenou. O novinku se hned začala zajímat média a oslavovat Teslu jako první auto bez řidiče, ačkoli sama společnost se snažila upozornit na omezení technologie. Zejména zdůrazňovala, že během celé jízdy se zapnutým *Autopilotem* by řidiči měli mít ruce na volantu. První smrtelná nehoda způsobená *Autopilotem* ve voze Tesla se stala 7. května 2016 na Floridě. Vůz narazil do přívěsu kamionu přejíždějícího na dálnici. Nehoda měla dvě příčiny: senzory vozu zmátl pohled na bílý kamion na světlé obloze, a pak i řidiče, který nedával pozor a nevěnoval se řízení (údajně se díval na film ze série Harry Potter). [Tegmark, 2017: 130]

Revoluce neuronových sítí dosáhla svého vrcholu 15. března 2016 kdy program AlphaGo porazil nejlepšího hráče Go na světě. Program vyvinula společnost DeepMind ze Spojeného Království, kterou v roce 2014 koupila společnost Google (přesněji holding Alphabet) za 650 milionů amerických dolarů. V době akvizice nebyly na webové stránce společnosti dostupné žádné bližší informace o technologiích a službách. Bylo zde však posláni společnosti – vyřešit *hádku intelligence*.

Jedním z prvních úspěchů společnosti bylo vytvoření programu, který v roce 2014 dokázal hrát klasické hry od ATARI, aniž by ho to někdo naučil. Max Tegmark popisuje moment, kdy program poprvé viděl v akci jako jeden z největších momentů, kdy mu “spadla čelist”. Napsat program, který dokáže hrát některou z jednoduchých her není zase nic tak složitého, programátoři z DeepMind však šli jinou cestou. Jejich cílem bylo vytvořit agenta využívajícího neuronovou síť, který se bude schopný naučit co možná nejvíce her. Síti neposkytli žádné informace o konkrétní hře ani popis vizuálních prvků, které se mohou objevit ani o vnitřním stavu použitého emulátoru. Program se učil pouze z vizuálního výstupu hry, odměn a výstupu terminálu. Tak, jak by se učil hrát lidský hráč. [Tegmark, 2017: 110-114]

Fungování lze vysvětlit na klasické hře Breakout, jejímž cílem je míčkem zničit všechny cihly v horní části obrazovky pomocí ovladatelné pohyblivé plošinky v dolní části. Umělé inteligenci bylo řečeno, aby maximalizovala skóre odesíláním čísel, která reprezentovala stisk jednotlivých kláves. Ze začátku program pouze bezradně jezdil plošinkou sem a tam a mýjel míček. Postupně začal zjišťovat, že trefování míčku vede k vyššímu skóre a s každou hrou ho trefoval častěji a častěji až do fáze, kdy ho byl schopný

trefit pokaždé nehledě na jeho rychlost. Nakonec, po 240 minutách hraní, se program naučil optimální strategii pro výhru. Vytvoření otvoru na straně a ponechání míčku zničit cihly zezhora což lze bez nadsázky považovat za inteligentní chování<sup>13</sup>. Jeden z programátorů se následně vyjádřil, že dokud tuto strategii neviděli u programu, tak ji neznali. [Wooldridge, 2021: 99-101]

DeepMind vyvinula jednoduchou ale opravdu schopnou technologii nazývanou *deep reinforcement learning*<sup>14</sup>. V případě hry Breakout to bylo vyšší skóre, pokud správně trefí míček. Vytvořili tak neuronovou síť, která předpovídala jaká sekvence čísel (tedy kláves) povedou k největšímu počtu bodů. Program pak vybral tu nejvýhodnější klávesu na základě aktuálního stavu hry. [ibid.]

Hra Go, ačkoliv má mnohem jednodušší pravidla než šachy, byl pro AI od začátku jejího vývoje obtížnou úlohou. Problémem je velikost hracího pole – deska má rozměry 19x19 políček, tedy 361 míst, kam lze položit hrací kameny. Šachy mají pole pouze 8x8 a tedy 64 míst pro figurky. Faktor větvení (průměrný počet tahů, které může hráč v daném okamžiku odehrát) je přibližně 250, u šachů je to přibližně pouze 35. Partie Go také mohou trvat dlouho, 150 tahů není nic výjimečného. AlphaGo používal dvě hluboké neuronové sítě, value network a policy network. Value network měla za cíl zabývat se výhradně odhadem, jak dobrá je aktuální pozice na hracím poli, zatímco policy network odhadovala, jaký bude nejlepší příští tah ve vztahu k aktuální pozici. Druhá síť byla trénovaná pomocí supervised learning za použití dat z odehraných her lidskými experty a následně pomocí deep reinforcement learning, kdy hrál program sám se sebou. Nakonec byly tyto sítě byly “propojeny” s pokročilou technikou prohledávání stavového prostoru metodou Monte Carlo. [Wooldridge, 2021: 101-103]

Před veřejným představením AlphaGo si společnost DeepMind najala evropského šampiona ve hře Go Fan Huiho, aby si proti programu zahrál. Program ho porazil ve všech pěti hrách a bylo to vůbec poprvé, co něco takového dokázal. Nedlouho po tom DeepMind oznámil zápas se světovým šampionem Lee Sedolem, který se odehrál 15. března 2016 v korejském Soulu – AlphaGo vyhrál 4:1. [Wooldridge, 2021: 101-102] Lee Sedol se po prohře omluvil veřejnosti, že selhal a rozhodl se ukončit profesionální kariéru. Reportérům po zápase řekl: “S nástupem umělé inteligence ve hrách Go jsem si uvědomil, že nejsem na

---

<sup>13</sup> Celý proces učení je velmi dobře vidět na Youtube videu “Google DeepMind’s Deep Q-learning Playing Atari Breakout” <https://tinyurl.com/atari.ai>.

<sup>14</sup> Viz kapitola technologie



vrcholu, i když se díky horečnému úsilí stanu jedničkou. I když se stanu jedničkou, existuje entita, kterou nelze porazit.”<sup>15</sup>. [Brown, 2019] V průběhu soutěže komentátoři několikrát poznamenali, že AlphaGo provádí zvláštní tahy, které by lidský hráč rozhodně neprovedl. Pokud máme snahu analyzovat hru z lidského hlediska a hledat motivy a tahy, které bychom dělali my dochází k antropomorfizaci. Snažit se takto pochopit program jako AlphaGo je zbytečné, takový program je optimalizován jen k jednomu účelu, a to hrát hru Go. Chceme mu přisoudit motiv, uvažování či strategii, v jeho neuronových sítích je však stejně jen dlouhý seznam čísel a mimořádné schopnosti jsou dány jen hodnotami vah. V současné době nemáme žádný způsob, jak tyto znalosti a jejich význam ukryté v hodnotách extrahovat nebo racionalizovat. [Wooldridge, 2021]

O rok později vydala společnost DeepMind aktualizovanou verzi AlphaGo Zero, která začínala s náhodnými neuronovými sítěmi a učila se pouze z her proti sobě. Během několika dní se program naučil hrát lépe než předchozí verze, kterou porazil a získal znalosti, které lidští hráči získávají roky. Vytvořil nové nekonvenční strategie a tahy. Na konci roku společnost představila jeho nástupce – program AlphaZero, který uměl hrát kromě Go ještě šachy a hru shogi. V průběhu následujících let se vývojářům například povedlo vytvořit i programy hrající RTS<sup>16</sup> *Starcraft II* lépe než 99.8 % lidských hráčů na platformě *Battle.net*. To vše pomocí stejného použití *deep reinforcement learning*. Posledním úspěchem do už tak neuvěřitelné řady je program *Player of Games* z roku 2021, který nově umí hrát i Poker a Scotland Yard. PoG pracuje pouze s jedním vysoce pokročilým algoritmem a evolučně se tak liší od všech předchozích. [ibid.]

V tento moment bych příběh umělé inteligence ukončila, respektive pouze tuto jeho kapitolu, protože celý příběh má otevřený konec. Poprvé v historii pravděpodobně máme technologie na to naplnit vize, které stály u vzniku celého oboru. Je pouze na nás jako na lidstvu, zda tuto příležitost využijeme k něčemu dobrému. Již v roce 2015 přední vědci z oblasti AI, robotiky a osobnosti spojené s technologiemi včetně Elona Muska, Stephen Hawkinga, spoluzakladatelů DeepMind či profesora Maxe Tegmarka podepsali otevřený dopis proti vývoji autonomních zbraní, které jsou velmi pravděpodobně reálné a šance, že by se dostaly do nesprávných rukou je poměrně velká. [Victor, 2015]

---

<sup>15</sup> „With the debut of AI in Go games, I’ve realized that I’m not at the top even if I become the number one through frantic efforts. Even if I become the number one, there is an entity that cannot be defeated.

<sup>16</sup> Real Time Strategy, druh počítačové hry

Ke zvážení by dle mého názoru také bylo upuštění od označení *umělá inteligence* – programy od AlphaGo ukázaly, že umí přemýšlet po svém a dosahovat lepších výsledků než lidé. V roce 2017 obletěla svět fantastická zpráva o chatbotech společnosti Facebook, kteří si vyvinuli “vlastní jazyk”, kterému již výzkumníci nerozuměli a raději experiment ukončili. Jakkoliv zpráva zněla pro laickou veřejnost děsivě, šlo pouze o běžný výzkumný (a úspěšný) experiment, u kterých je běžné jejich ukončení, pokud přestanou produkovat smysluplné výsledky. [Kučera, 2017] Jako poslední argument pro upuštění od tohoto označení bych chtěla uvést, že ani po skoro 66 letech od letního semináře v Dartmouth College neexistuje obecně přijímaná definice, co vlastně termín znamená. Upustit od něj by nám, jako společnosti mohlo otevřít nové obzory, v tomto příběhu jsme koneckonců viděli, že několikrát došlo k přejmenování různých technologií, aby se zbavili zbytečných, zakořeněných předsudků. Osobně se mi líbí vize inteligentních agentů, kteří nám pomáhají v každodenním životě, aniž bych musela přemýšlet o tom, zda vykazují prvky lidské inteligence.

## 1.6 Technologie AI

### 1.6.1 Expertní systémy

V kontextu umělé inteligence označujeme jako expertní systém aplikaci, která má napodobovat rozhodovací schopnosti lidského experta a umožnit tak uživateli konzultaci problému. Tyto aplikace jsou navrženy tak, aby řešily složité problémy, kde je velká míra neurčitostí a obsahují znalostní bázi vytvořenou již zmíněnými experty. První takové systémy začaly vznikat v 70. letech 20 století a k většímu rozšíření došlo v letech 80. letech. [Leones, 2002: 1-22] Expertní systémy lze považovat mezi jednu z prvních skutečně úspěšných aplikací umělé inteligence v softwaru. [Russel, Norvig, 2022: 71-73]

První expertní systém byl vyvinut v roce 1974 pod názvem MYCIN a jeho úkolem bylo doporučování vhodných léků na základě analýzy zdravotní karty pacienta a zohledňovat při tomto výběru již užívané léky. Dalším obecně známým expertním systémem je CLIPS, který vyvíjela mezi roky 1985-1996 americká organizace NASA. Po ukončení vývoje byl program vydán jako public domain<sup>17</sup> včetně zdrojových kódů. Projekt je stále aktivně spravován a kdokoliv si ho může vyzkoušet. [Nillson, 2009]

---

<sup>17</sup> Public Domain

Některé jejich koncepty používají i moderní aplikace, avšak jsou implementovány mnohem jednodušším způsobem a nejsou považovány za AI.

### 1.6.2 Machine learning (strojové učení)

Strojové učení je základním stavebním kamenem velké části současných aplikací AI. Ačkoliv se termín *strojové učení* někdy zaměňuje s pojmem AI, není to správně. Strojového učení je pouze jedna z jeho podmnožin. Například expertní systémy nebo znalostní grafy strojové učení nepoužívají, a přesto je považujeme za AI. To, co odděluje machine learning je zejména schopnost učit se a modifikovat svůj chod na základě vložených dat. Díky tomu je program méně závislý na lidských expertech a dokáže zlepšit svou výkonnost bez vnějšího zásahu. [Mitchell, 1997]

Tom. M. Mitchell ve své knize *Machine Learning* [Mitchell, 1997: 3] definuje strojové učení jako: “A computer program is said to learn from experience  $E$  with respect to some class of tasks  $T$  and performance measure  $P$ , if its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ .”

Tuto na první pohled poněkud složitou definici lze vysvětlit na příkladu hraní dámy, který sám autor ve výše zmíněné knize používá. Pokud se program učí hrát dámu, může se zlepšovat svůj výkon (měřený jako počet výher) pomocí zkušeností získaných z her, které hraje sám ze sebou. Obecně lze říct, že aby se program mohl efektivně učit, musí mít definované tyto tři vlastnosti: úlohu, výkonnostní měřítko a zdroj těchto zkušeností. V případě dámy tedy platí:

- Úloha  $T$ : hraní dámy.
- Výkonnostní měřítko  $P$ : počet vyhraných her proti soupeřům.
- Zdroj zkušeností  $E$ : hraní cvičných her sám proti sobě.

Zjednodušeně řečeno je strojové učení proces, podle kterého program postupuje, aby dosáhl umělé inteligence. Používá různé algoritmy k nalezení vzorů ve vstupních datech a snaží se je používat k vytvoření datového modelu za účelem predikování výsledku. Předpokladem pro kvalitní výstup je tedy kvalita vstupních dat a počet opakování. V konečném důsledku jde tedy jen o velice sofistikovaný algoritmus. [Nilsson, 2009: 495]

### 1.6.2.1 Neuronové sítě

Základním stavebním kamenem většiny moderních systémů AI jsou neuronové sítě. Jsou jedním z hlavních trendů již od 60. let 20. století a v posledních letech došlo k jejich velkému zlepšení zejména díky výkonnějším počítačům a technickým pokrokům obecně. Jejich koncept lze využít ve velkém množství aplikací jako je zpracovávání přirozeného jazyka, kategorizace grafiky, ale i při strojovém hraní her. Je nutné podotknout, že neuronová síť vždy funguje na základě pravděpodobnosti a neposkytuje exaktní výsledky. [Wooldridge, 2021: 92]

Neuronová síť je systém výpočtů inspirovaný tím, jak fungují biologické neuronové sítě. V lidském mozku jsou jednoduché výpočetní jednotky (neurony) které pracují paralelně a vzájemně spolu komunikují. Umělý neuron je jednoduchá výpočetní jednotka s matematickou funkcí, která zpracovává vstup, komunikuje s ostatními podobnými jednotkami a produkuje výstup. Vstupem neuronu může být informace z vnějšku či z jiného neuronu, každý tento vstup je opatřen váhou. Po přijmutí vstupů jsou hodnoty vynásobeny váhami a pokud výsledek překročí stanovený práh, dojde k jeho transformaci předem danou funkcí a je odeslán na výstup. [ibid.] Viz Obrázek 2

Jak proběhne transformace vstupů na výstupy v neuronové síti závisí tedy na hodnotě vah a prahů. Tyto hodnoty se stanoví v procesu, pro které bylo zvoleno označení učení. V případě neuronových sítí rozlišujeme dva druhy učení – s učitelem a bez učitele. Při učení s učitelem máme známou sadu dat a požadovaný výstup. Pokud po spuštění neuronové sítě nezískáme požadovaný výsledek dojde k upravení vah a prahů a pokus opakujeme, dokud síť nedosahuje chtěného výsledku. [Nillson, 2009: 507-515]

Dojde-li ke správnému nastavení všech hodnot a síť je správně navržena, je schopna řešit i problémy se kterými se během učení nesetkala. Aplikaci lze již považovat za jakousi umělou inteligenci. Při učení bez učitele nemáme určené výstupy konkrétní datové sady. Toho se nejčastěji využívá při shlukové analýze, při které chceme vstupní data rozdělit do konečného množství tříd. [ibid.]

### 1.6.2.2 Deep Learning

Deep learning (hluboké učení) je speciální podmnožinou strojového učení. Pokud je tento termín použit, nejčastěji odkazuje na hluboké neuronové sítě (deep artificial neural networks). Termín “hluboké” určuje počet vrstev v neuronové síti – mělké neuronové sítě mají jednu tzv. skrytou vrstvu oproti hlubokým, které mají těchto vrstev více. [Wooldridge,

2021: 95] Romportl technologii popisuje následovně [Romportl, 2022]: „Je to v podstatě vysoce paralelní výpočetní proces popsáný velkým grafovým výpočetním schématem, který umožňuje provádět značné množství malých výpočtů nad mnohadimenzionálními vstupními daty, a tím jejich dimenzionalitu postupně redukovat.“

Oproti klasickému strojovému učení se liší především tím, jak se jednotlivé algoritmy učí a jaké množství dat používají. Probíhá zde více automatizace při extrakci prvků a dochází tak k odstranění potřeby některých zásahů člověka. V případě klasického strojového učení musí experti ručně zadat charakteristiky a rysy vstupů tak, aby vznikly rozlišovací vzory (váhy a prahy). Program využívající hluboké učení je schopný si vytvořit tyto vzory sám a třídít podle nich výsledky. Zjednodušeně řečeno stačí programu posílat data a říkat mu, že je výsledek špatný, dokud nebude správný. [Wooldridge, 2021: 95]

Další výhodou je schopnost zpracování mnohem většího množství dat, a to i těch nestrukturovaných, kterých se vytváří v praxi nejvíce. Na druhou stranu toto může být i nevýhodou, protože velké množství dat je nutným předpokladem k dobrým výsledkům. Klasické strojové učení si vystačí s méně daty, pokud známe výsledky. [Nillson, 2009]

V současné době často splývá označení deep learning s neuronovými sítěmi a pokud je někde zmíněno, ve velké části případů se hovoří právě o něm.

### 1.6.2.3 Reinforcement learning (RL)

Při reinforcement learning (posilovací učení) nejsou programu poskytnuta tréninková data. Program zkouší různé strategie a na základě výstupu je mu poskytnuta zpětná vazba, zda jeho strategie byla dobrá nebo špatná. Celý koncept je založen na behavioristické psychologii, získání pozitivní odměny zvyšuje vůli něco udělat znovu a naopak. Někdy se také opisuje jako učení se z chyb. [Nillson, 2009: 515-526]

Velmi často se používá k trénování programů pro hraní her, na kterých lze i snadno demonstrovat. Program hraje hru a jeho zpětná vazba je ukazatel skóre. Pokud ji hraje dobře, skóre se zvyšuje (pozitivní zpětná vazba), pokud však hraje špatně, skóre zůstane stejné (negativní zpětná vazba). Zpětná vazba, kterou dostane, ať už pozitivní, nebo negativní, je v tomto případě odměna. Odměnu program zohlední při příští hře. Pozitivní odměna povede ke hře stejným způsobem a negativní odměna povede ke změně strategie. [ibid.]

Klíčovým problémem při použití reinforcement learning je, že v mnoha situacích může odměna přijít až za dlouhou dobu. Pro program je obtížné zjistit, které akce byly dobré a které špatné. Pokud program hraje například šachy, pravděpodobně všechny tahy ve hře nebyly špatné. Tento postup učení se také nehodí na příliš jednoduché problémy.

#### 1.6.2.4 Deep reinforcement learning (DRL)

Nejnovější a pravděpodobně nejefektivnější technika strojového učení se nazývá deep reinforcement learning (hluboké posilovací učení). Jeho první funkční implementaci představila firma DeepMind v roce 2016. Jedná se o kombinaci technik deep learning a reinforcement learning. Technika obsahuje výhody obou zmíněných, zvládne zpracovat nestrukturovaná data pomocí neuronové sítě a machine learning přináší dovednost rozhodování a hledání nejlepší strategie. [Deepmind, 2016]

Ve veřejném prostoru se nejčastěji setkáme se zprávami, že DRL vyhrává nad lidskými protivníky v počítačových hrách. Díky DRL byl například naučen i program AlphaGo, který porazil mistra ve hře Go. Technika se používá i v autonomních vozech nebo v robotice. Lze říci, že využití díky své všestrannosti našla ve všech odvětvích podnikání. [Torres, 2020]

#### 1.6.2.5 Generative adversarial network (GAN)

Generative adversarial network (generativní kontradiktorní síť) je typ strojového učení, který navrhl se svým týmem Ian Goodfellow v roce 2014. Je založena na principu dvou soupeřících neuronových sítích ve hře s nulovým součtem (zisk jedné znamená ztrátu druhé). [Brownlee, 2019]

Souboj lze přirovnat ke vztahu mezi padělatelem obrazů a detektivem, kteří se opakovaně snaží přelstít jeden druhého. Obě sítě jsou naučeny na stejném souboru dat. První z nich, známá jako *generátor*, má za úkol vytvářet umělé výstupy, jako jsou fotografie nebo rukopis, které jsou co nejrealističtější. Druhá síť, známá jako *diskriminátor*, je porovnává s pravými obrázky z původní datové sady a snaží se určit, které jsou pravé a které falešné. Na základě těchto výsledků generátor upraví své parametry pro vytváření nových obrazů. A tak to pokračuje, dokud diskriminátor nedokáže rozlišit, co je pravé a co falešné. [ibid.]

Využití GAN je opravdu různorodé, lze díky němu například zvyšovat rozlišení obrázků (textury počítačových her [Barnett, 2019]) a vytvářet umělecká díla [Vincent, 2019], a to klidně nekonečně. Technologie však není omezena pouze na obrázky, lze vytvářet i hudbu. [Shiledarbaxi, 2021]

#### 1.6.3 Znalostní grafy

Znalostní grafy jsou často opomíjenou technologií, ačkoliv se s jejich implementací na vyhledávači Google každodenně setkávají miliony uživatelů. V roce 2014, kdy

technologii Google představil, byl jejich znalostní graf synonymem pro technologii, dnes ho používají i jiné velké společnosti jako Apple, Amazon a Microsoft. [Morrison, 2022]

Jde o znalostní bázi, která k uložení dat používá datový model nebo topologii s grafovou strukturou. Znalostní grafy se často používají k ukládání vzájemně propojených popisů různých entit (objektů, událostí, situací nebo abstraktních pojmů) a popisu jejich vztahů (sémantiku).

Od roku 2018 zařazuje společnost Gartner znalostní grafy do svých předpovědí budoucích trendů a předpovídá největší rozšíření v horizontu 5 až 10 let. [Gartner, 2021]

## 1.7 Současnost

### 1.7.1 Zpracování textu

Zpracování přirozeného textu se posunulo kupředu zejména díky novým jazykovým modelům využívajícím pokročilé neuronové sítě. Modely jako GPT, ELMo a BERT se dokáží naučit v jakém kontextu se daná slova používají včetně gramatiky a jejich vztahu k okolnímu světu.

Nejnovější iterace nejznámějšího modelu neziskové organizace OpenAI je GPT-3. Model byl naučen pomocí 700 gigabitů (přes bilion slov) textových dat z celého internetu, včetně Wikipedie a sbírky digitalizovaných knih. Inteligenci model získává díky tzv. bottom up přístupu – předpovídá další slovo, které by mohlo s největší pravděpodobností následovat, tímto stylem pak sestaví celou větu. Učení modelu GPT probíhá pomocí “výzvy”, zadáním několika odstavců či vět ze známého textu a požádáním modelu, aby navrhl seznam slov, která by mohla následovat seřazených podle pravděpodobnosti. Ze začátku půjde pravděpodobně o nesmysly, avšak někde se bude vyskytovat slovo, které dává smysl. Model pak přiřadí parametrům, které vedly k tomuto správnému slovu vyšší váhu, a naopak oslabí ty cesty, které nikam nevedly. Při dostatečném počtu interakcí se model naučí, jak sestavovat smysluplné věty. V současné době dokáže tímto postupem model vytvořit i delší pasáže textu, které nelze rozeznat od těch vytvořených člověkem, včetně více či méně povedených básní nebo dokonce zdrojového kódu programu. Jeden z nejvíce fascinujících důkazů je však schopnost modelu rozeznat textový vstup uživatele v přirozeném jazyce a na jeho základě vyprodukovat nějaký výstup. [Anderson, 2021]

Ačkoliv už možnosti modelu GPT-3 vypadají jako z oblasti sci-fi, společnost Google se s Pathways Language Model (PaLM) představeným v dubnu tohoto roku (2022) rozhodla jít ještě dále. Pro porovnání velikosti, model GPT-3 175 000 milionů parametrů, zatímco

PaLM disponuje 540 000 miliony. Díky tomu model dokáže velmi přesně odpovídat na otázky a dokáže a uvažovat. Vysvětlení toho, jak neuronová síť dospěla ke svému závěru bylo donedávna brané jako její největší slabina, možnosti modelu PaLM vyřešit slovní úlohu a vysvětlit v přirozeném jazyce, jak dospěl k výsledku jen dokazuje vyspělost dnešních technologií. Společnost také ukázala příklady, kdy model byl schopný vysvětlit vtíp nebo opravit zdrojový kód tak, aby ho bylo možné zkompilovat. Schopnosti, které ještě před několika lety byly nemožné. [Narang, Chowdhery, Brynjolfsson, 2022] Viz Obrázek 13

Z pokroků ve zpracování textu benefituje celá řada aplikací jako je strojový překlad, klasifikace textu či rozpoznání řeči. To všechno vede k lepším interakcím mezi člověkem a umělou inteligencí, například hlasovým asistentem. Jednou z výzev je to, jak zajistit kvalitní data pro modely, které jsou využity pro proces učení. S tím, jak rostou jejich schopnosti a velikost, roste i potenciální předpojatost. Dle výzkumu Stanfordské univerzity je model s 280 miliardami parametrů o 29 % toxičtější (obsahující urážky, obscénosti, výhrůžky apod.), nežli nejpokročilejší model z roku 2018 s 177 miliony parametrů.

Když vidíme, jaké možnosti umožňují současné jazykové modely, nabízí se otázka, zda jsme na ně jako společnost připraveni. V nedávné době obletěla svět zpráva o výzkumníkovi společnosti Google, který při experimentu dospěl k závěru, že program nabyt vědomí. [Kasík, 2022] Pokud se toto může stát expertovi na vysoké úrovni, jak na podobné technologie pak bude reagovat laická veřejnost? Vědomí programu bylo samozřejmě vyvráceno. Tomáš Mikolov z Českého institutu informatiky, robotiky a kybernetiky ČVUT vysvětluje: „Modely si dokážou zapamatovat miliardy vět z internetu, které napsali lidé. A tyto věty pak, zjednodušeně řečeno, ve vhodných kontextech opakovat. Lidé se tak vlastně dívají do zrcadla. A někteří mají dojem, že je ten obraz živý.“. Umělá inteligence dokáže generovat poměrně zajímavé básně proměnlivé kvality. Viz Obrázek 8 a Obrázek 9

### 1.7.2 Zpracování obrazu

Pokročilé zpracování obrazu je nyní integrované v celé řadě aplikací. Od změny pozadí během videokonference po vytváření falešných fotorealistických obrazů známých jako *deep fakes*. [Shen, 2018] Přístupy často využívají *deep learning* k rozpoznávání objektů, klasifikaci fotek nebo rozpoznání textu.

Díky výkonnějšímu hardwaru se zkrátila i doba učení. Programy využívající síť ImageNet, která obsahuje přes 14 milionů tréninkových dat, dokončí proces 100 krát rychleji



než před 3 lety [Zhang, 2021: 49] Trend můžeme pozorovat i při licencování softwaru, spousta autorů zveřejňuje své zdrojové kódy a dává je k dispozici komunitě.

Jeden z takových je systém pro detekci objektů v reálném čase YOLO (You Only Look Once). Ačkoliv je stále k dispozici a vznikají nové projekty s jeho použitím, vývojář Joseph Redmond se rozhodl ukončit jeho oficiální vývoj kvůli dopadu své práce. Dle svých slov již nemohl dále “ignorovat vojenské aplikace a obavy o ochranu soukromí”. [Redmond, 2020]

Významným úspěchem bylo vytvoření technologie GAN<sup>18</sup>, umožňující téměř nekonečné možnosti manipulace s fotografiemi, ale i videem [TZABAN, 2022] Dříve bylo nutné k takovým úpravám mít specializovaný software a znalosti, AI podstatně snížila vstupní bariéry. To sebou bohužel nese i negativa ve formě různých podvodů, kdy útočník může vytvořit falešnou identitu k získání přístupu k nějaké službě či vytvořit podvrhnutý sexuální obsah a použít ho jako způsob nátlaku. Jedním z trendů je tedy vytvořit nástroje, jak tyto stále dokonalejší podvody odhalit. [Groh, Epstein, Firestone a kol, 2021] Sociální rovině tohoto problému se budu nadále věnovat v další části práce.

Umění vytvořené pomocí GAN se také dostalo do aukčního sálu. V roce 2018 byl prodán portrét fiktivního Edmonda Belamyho za 432 500 amerických dolarů (přes 10 milionu korun) v aukčním domě Christie’s [Christie’s, 2018] Zajímavostí je, že kromě samotného Edmona byla vytvořena celá rodina stejným způsobem.

### 1.7.3 Robotika a mobilita

V posledních letech přispěla ke zlepšení v oblastech robotiky zejména větší sofistikovanost senzorů, více komunikačních možností a celkově vývoj v široké řadě technologií. Roboti stále nemohou využít všech pokroků na poli umělé inteligence zejména kvůli fyzikálním omezením.

V průmyslové sféře se stále více uplatňuje deep learning a předpokládá se, že má velký potenciál iniciovat v budoucnu rapidní změny. Pokrokové firmy již teď používají deep learning a reinforcement learning ke změně způsobu ovládání strojů v továrnách a skladech. Řeší obtížné úkoly, například jak přimět roboty, aby detekovali a vybíraly předměty různých velikostí a tvarů ze zásobníků. Způsoby, jakými vyrábíme a přepravujeme zboží, jsou do značné míry závislé na spolupráci skupin strojů a efektivita a odolnost těchto strojů jsou

---

<sup>18</sup> Viz kapitola GAN

základem naší ekonomiky a společnosti. Pokud by se podařilo tuto technologii zavést do praxe, jistě by znamenala výraznou změnu trhu práce.

Autonomní vozidla a různé techniky samořídících vozidel se stále více dostávají do běžného komerčního sektoru. Tento vývoj ovlivňuje celý automobilový průmysl, ale také městské plánování. Ve stále více sériových vozech můžeme nalézt asistenční systémy využívající pokročilé senzory a umělou inteligenci k parkování, udržování vozidla v jízdním pruhu nebo k dodržování bezpečné vzdálenosti.

Společnost Waymo, spadající pod holding Alphabet, začala na konci března 2022 testovat své plně autonomní taxi v ulicích San Franciska a doufá v co nejbližší získání povolení pro provoz bez řidiče. [Ohnsman, 2022] V Evropě je průkopníky autonomní mobility pravděpodobně Německé město Monheim nad Rýnem, kde od začátku roku 2020 provozují flotilu 5 autonomních elektrických mikrobuseů. Mikrobus jezdí na předem definované trase v běžném provozu. [Marquardt, 2020] O rok později začala testovat podobný koncept i Francie, konkrétně ve školním kampusu ve městě Toulouse. [Bateman, 2021] Na asijském kontinentu můžeme nalézt autonomní taxi v Pekingu. Problém kontroly je zde novátorsky řešen pomocí 5G sítě, která umožňuje v případě problémů převzít na dálku kontrolu nad vozem. [Templeton, 2021]

Optimistické předpovědi o létajících a autonomních vozech se zatím nenaplnily. Důvodem je nejen potřeba mimořádné úrovně bezpečnosti ve složitém prostředí (která se ukázala mnohem náročnějším a nákladnějším, než se původně předpokládalo [Gessner, 2021]), ale také legislativa. V té má momentálně nejlépe nakročeno Německo. [Ewing, 2021]

#### **1.7.4 Zdravotnictví**

Využívání AI v biomedicínských aplikacích rok od roku roste – zejména v diagnostice a při objevování nových léků. Technologie se přesunuly z akademické sféry do té komerční. Pokročilé nástroje pro odhalení rakoviny prsu jsou dostupné už i v České republice. [Kuklová, 2022] Kromě léčby umělá inteligence podporuje celou řadu dalších procesů ve zdravotnictví, například pomáhá předpovídat délku trvání operací a optimalizovat tak plánování. [Bajwa, Munir, Nori a kol., 2021]

Velký potenciál v biochemii a medicínských aplikacích má průlom, za kterým stojí společnost DeepMind. V roce 2020 vytvořili vývojáři neuronovou síť, která dokáže předvídat tvar bílkovin a vyřešila tak “problém skládání proteinů”. Tento problém se věda

snažila bezvýsledně vyřešit 50 let. Sami vývojáři to označují za jeden z jejich zatím nejvýznamnějších úspěchů. [Deepmind, 2020]

AI hrála také velkou roli ve vývoji vakcíny proti Covidu 19. Díky pokročilým technologiím byly klinické testy nové vakcíny zahájeny již 3 měsíce po prvních případech nákazy. Vnější bílkoviny viru obsahují desítky tisíc dílčích složek. Modely s využitím machine learningu dokázaly rozřadit a předpovědět, které dílčí složky jsou nejvíce imunogenní. Tedy schopné vyvolat imunitní reakci, a tím nasměrovat výzkumníky při navrhování cílených vakcín. Využití AI při vývoji vakcín může v budoucnu způsobit revoluci ve způsobu výroby vakcín obecně. [Park, 2021]

Všechny pokroky v oblasti zdravotnictví a přírodních věd nám dávají další jistotu, že se umělá inteligence stane jedním z nejužitečnějších nástrojů lidstva při rozšiřování hranic vědeckého poznání. I v této oblasti jsou s AI spojena etická rizika, která budou rozvedena v analytické části s konkrétními případy z posledních let.

### **1.7.5 Finance**

Vliv umělé inteligence na finanční sektor je od roku větší. Nejvíce viditelní jsou bankovní asistenti, které má dnes snad již každý bankovní ústav i v České republice. Ten společnosti AirBank je dokonce postaven na legendárním Watsonovi od IBM. Díky těmto asistentům mají zákazníci určitý stupeň podpory v jakoukoliv denní či noční hodinu a nejsou odkázáni na práci operátorů či přepážek. [IBM, 2021]

Klasickým využitím AI je detekce podvodů s platebními kartami a platbami obecně, díky IBM je to i technicky zajímavé. Mainframy IBM hrají ve finančním světě velkou roli. Společnost uvedla, že 45 z 50 největších světových bank používá její infrastrukturu, která zpracovává 70 % celosvětových transakcí co se hodnoty týče. Během pandemie Covidu-19 navíc počet podvodů při platbách na internetu ještě narostl. Klasické modely hlubokého učení dokáží zlepšit v reálném čase detekci podvodů zhruba o 60 %, je to však vykoupené delší odezvou. Propustnost takového systému je 1000 až 1500 transakcí za sekundu s odezvou mezi 50 a 80 milisekundami. Server osazený 32 novými procesory IBM Telum a spuštěný v dubnu tohoto roku (2022) je schopný za sekundu zpracovat přes 100 000 transakcí s odezvou okolo 1 milisekundy. Tento značný nárůst výkonu umožní klientům kontrolovat každou transakci, a nikoliv jen část, jak tomu bylo dříve. [Groenfeldt, 2022]

Velkou příležitostí k využití technologií AI představuje compliance, nicméně podle zprávy společnosti Deloitte pouze 4 % respondentů ve finančních službách ji k tomuto účelu používá. Více než polovina (61 %) dotazovaných však očekává výraznou změnu jejich

odvětví díky umělé inteligence v příštích třech letech. Téměř všichni (95 %) dotázaní vyjádřili obavy z etických rizik pramenících z použití AI ve finančním sektoru. [Niekerk, 2022]

Stále kontroverzní zůstává způsob, jak AI rozhoduje o půjčkách a předpojatost s tím spojená. Tomuto problému se budu více věnovat v analytické části práce.

### 1.7.6 Cesta k obecné inteligenci

Rok 2022 se zapíše do historie AI stejně jako rok, kdy Deep Blue porazil Garyho Kasparova, nebo kdy AlphaGo poprvé porazilo šampiona v Go. Společnost DeepMind 12. května představila *všeobecného agenta* Gato. Gato zvládá velkou škálu úkolů od vytváření poezie, hraní her Atari, anotaci obrázků a po připojení robotické paže i skládání kostek. Revoluční je to, že všech přes 600 činností dokáže najednou, aniž by se musel znovu vytrénovat. [Deepmind, 2022]

Ředitele výzkumu společnosti DeepMind, Dr Nando De Freitas, na svůj Twitterový účet napsal: “Ted’ už jde jen o měřítko! Hra skončila! Jde o to, aby tyto modely byly větší, bezpečnější, výpočetně efektivnější, rychlejší při vzorkování, inteligentnější paměť, více modalit, inovativní data, on/offline... Řešení těchto výzev je to, co přinese AGI.”. [Cuthbertson, 2021] Futurista Tristan Greene jeho názor nesdílí. Pokazuje na to, že ačkoliv je to, co Gato umí úctyhodné, není o moc blíže obecné inteligence než hlasový asistent Alexa. Nejedná se o obecnou umělou inteligenci, ale o několik naučených *úzkých* modelů, které jsou úhledně zabaleny do jedné aplikace. [Greene, 2022]

Společnost DeepMind pracuje na AGI již více než deset let a OpenAI od roku 2015. A ani jeden z nich nebyl za tu dobu schopen vyřešit úplně první problém na cestě k vyřešení AGI: vytvořit umělou inteligenci, která se dokáže učit nové věci bez tréninku. Dle Greena by bylo nasnadě přehodnotit časové plány ohledně vzniku AGI, za předpokladu, že se tak nestane nějakou šťastnou náhodou. [ibid.] “Neřekl bych "nikdy", protože to je jedno z mála prokletých slov vědy. Ale vypadá to, že AGI se v našich životech neobjeví.” [Greene, 2022]

K podobnému názoru dospěl i Tegmark [Tegmark, 2017:171] Tvrdí, že už neexistuje pádný argument, že nám chybí dostatečná hardwarová síla nebo že to bude příliš drahé. Dokládá to výpočtem Hanse Moravce ohledně potřebného výkonu na emulaci mozku. Moravec dospěl k závěru, že stejný výkon poskytne optimalizovaný počítač z roku 2015 v ceně 1000 amerických dolarů [Tegmark, 2017:170]

Tegmark zároveň možnost AGI, která dosáhne lidské úrovně (nebo ji přesáhne) úplně nezavrhuje. Nikdo nedokáže říci, jak daleko jsme od pomyslné cílové rovinky. Algoritmy,

architektury a software se neustále zlepšují. Do oboru také nastoupilo mnoho talentů, což dokazuje i úspěch v podobě agenta Gato.

### 1.7.7 Shrnutí

Jak vyplývá z předchozí část, cesta k superinteligenci může být ještě daleká. Proto jsem se pro účely této práce rozhodla zaměřit se na současné problémy spojené s umělou inteligencí a to:

1. Obecná nejasnost, která kolem pojmu AI panuje. A to jak z technického hlediska, kdy neexistuje jedna definice a kdy to, co bylo ještě před několika lety považováno za AI je dnes pouhou aplikací matematiky. Ale také z nejasných představ lidí o ní díky vysoké bariéře expertního vědění kdy až do nedávné doby laik čerpal informace pouze médií a sifí.
2. *Code bias* neboli předpojatost kódu, kdy algoritmy AI jsou založeny na starých datech.
3. Schopnost AI přesvědčivě vytvářet texty a obrázky a tím manipulovat veřejné mínění.

## 2 Metodologie

Pro účely mé diplomové práce zabývající se otázkou, zda společenské vědy dokáží anticipovat budoucnost v případě umělé inteligence, a tedy i relevance sociálních věd pro tento obor se nabízela řada metod k analýze psaného textu. Byla jimi například frekvenční analýza nebo diskurzivní analýza. Diskurzivní analýzu jsem zavrhla, jelikož mne zajímalo naplnění nepřímých řečených vizí sociálních vědců a ne přímo rozdíly v diskurzu obou vědních disciplín. Frekvenční analýza se pro hledání kontextů mezi společenskovední literaturou a tou zabývající se umělou inteligencí jevila jako lepší volba, ale i tu jsem pro účely mé práce vyhodnotila jako nevhodnou z důvodu rozdílných systémů pojmů. Ty se liší nejen mezi oběma odvětvími, kde se například v technicky zaměřené literatuře objevuje řada anglikanismů, ale i v rámci společenskovední literatury, kdy každý autor používá jiné pojmosloví nebo dokonce přichází s vlastním. Právě díky hledání souvislostí a smyslů v textu často nevyřčených jsem se po konzultacích s vedoucím práce rozhodla pro metodu Althusserova symptomálního čtení a prvky hermeneutiky.

### 2.1 Althusserovo symptomální čtení

Metoda symptomálního čtení je jednou z metod hlubšího porozumění textu, kdy se autor vlastnímu zkoumanému předmětu nevěnuje ať už z důvodu působení v jiné oblasti nebo by se, jako v případě této práce, z důvodu tvorby v jiné době aktuálními otázkami ani nemohl zabývat. Jejím autorem je francouzský filosof Louis Pierre Althusser, který patří k nejvýznamnějším marxistickým teoretikům dvacátého století. Právě filosofická otázka jak skutečně pochopit Marxe stála i za objevením metody symptomálního čtení.

Althusser si všiml, jak Marx, aby pochopil podstatu věcí a světa, četl texty způsobem, kdy v jejich konkrétní podobě četl i jejich abstraktní podstatu. [Althusser, 1970] Podle něj tuto metodou objevil nejprve Marx, který ji však teoreticky nevypracoval, ale je v jeho díle aplikovaná a bylo ji nutné zpracovat. Toho se zhostil právě Althusser. [Kužel, 2014: 138] Popsal takzvané *bezprostřední čtení*, kdy postupně odhalujeme obsah a na konci objevíme pravý význam. Takové čtení má podle něj původ v náboženství a touze objevit skrytý význam čekající na odhalení. [Althusser, 1970] Tento typ čtení v *Kapitálu* mizí, ale objevují se v něm dva jiné typy.

První nazývá Althusser *retrospektivní teoretické čtení*, kdy Marx čte klasické ekonomy skrze jejich diskurz. Při něm nachází i mezery a problémy, které například Smith přehlédl. Ty jsou ale způsobené pouze přehlédnutím jedince, ale nemění diskurz. Althusser

u Marxe nachází i druhý typ čtení, *čtení symptomální*. [Kužel, 2014: 65] Díky takovému čtení je možné objevit i rovinu nevědomí. Právě díky němu je možné hledat otázky, které autor v díle nezamýšlel. To ilustroval na práci Marxe s mezerami a prázdnými místy v textech autorů klasické politické ekonomie. Jsou to místa pro ekonomy neviditelná, ačkoliv je mají před očima. Není to ale o individuální nedostatečnosti autorů, ale o jejich politickém vidění a zároveň nevidění. Althusser ale odrazuje od prostorových metafor, jelikož teoretické pole není ohraničeno a za jeho hranici se nenachází jiné. Naopak druhé pole, které vzniklo vyloučením všeho, co neodpovídalo limitům teoretického pole, se nachází uvnitř. [Hauser, 2011]

Při teoretické činnosti probíhající ve viditelném poli a pod vědomou kontrolou nepozorovaně vzniká něco nového díky přehlédnutí nebo vynechání. Althusser tato prázdná místa nazývá teoretickými symptomy. Tyto symptomy ukazují, že se v rámci viditelného sekundárně vyprodukovala i nová problematika vůči které je ale daný systém vědění slepý. Nový text je pak tedy výsledkem změny struktury teoretického pole. [ibid.]

Althusser nazývá tento způsob jako symptomální, jelikož změna struktury nového textu se dala vyzorovat z daných symptomů. Tento způsob čtení odlišuje od komparativního čtení, jelikož během srovnávání čteme první text a na něj přiložíme druhý a pouze konstatujeme, kde se shoduje, kde liší a kde má chyby a mezery. Symptomální čtení nebere předchozí jako nedostatky, ale jako symptomy nové problematiky. Druhý text pak používá ve vztahu k těmto symptomům. Jde tak o čtení neviditelného latentního obsahu. [ibid.]

## 2.2 Hermeneutika

Pojem hermeneutika jako věda má dlouhou historii, kdy se používala k výkladu náboženských textů a během let se vyvíjela. Podle Grondina byl tento pojem nejprve používán jako umění výkladu. Její záměr byl normativního rázu a omezovala se na to dát interpretativním vědám metodické pokyny, aby se zabránilo svévoli při interpretaci. [Grondin, 1997: 13] V šedesátých letech dvacátého století byla hermeneutika začleněna do sociálních věd jako metoda výzkumné práce. V moderní době byla hermeneutika rozpracována představiteli německé filosofie jako byl například Martin Heidegger, Georg Simmel a další. Samotný původ slova hermeneutika je však nejasný. Lze ho odvodit od Herma, posla bohů, kterému je v řecké mytologii přisouzen objev jazyka. Druhé vysvětlení může být ve slovesném základu hermeneuein, tedy říkat nebo vyslovit. [Hendl, 2005: 73]

Základem hermeneutiky je kruhový pohyb výkladu v takzvaném hermeneutickém kruhu a tedy probíhá cyklicky. Proces začíná prvním předporozuměním, kdy máme prvotní znalost textu, které tvoří základ pro jednotlivé analýzy. Výsledky těchto analýz pomohou zpětně pochopit celý text. [ibid.]

Hendl shrnuje v devíti krocích základní kroky textové hermeneutiky:

1. Je nutné adekvátně přiřadit otázku nebo specifikovaný problém k celku textu.
2. V textu určujeme místa, která představují odpověď na položenou otázku.
3. Přistupujeme k procesu dokazování, přičemž využíváme citáty textu.
4. Pracujeme s primárními nebo sekundárními texty, jež mohou odkazovat k dalším autorům. Spíše používáme primární texty.
5. Přistupujeme k vyložení imanentního obsahu textu, který se nejdříve vztahuje k výpovědím daného autora, zpracováváme také latentní témata.
6. Interpretace „posilujeme“ a verifikujeme ostatními textovými zdroji stejného autora.
7. V rámci komparativního postupu porovnáváme interpretace textů různých autorů, jež mají vztah ke stejnému problému.
8. Je nutné zodpovědět otázku spolehlivosti textů (bezrozporná argumentace, původ textů, původnost a úplnost atd.).
9. S interpretací spojujeme i vysvětlení, tzn. hledáme odpovědi na otázku: „Proč se tak děje?“ [Hendl, 2005: 74]

## 2.3 Výběr literatury k analýze

Při výběru textů pro analýzu jsem nejdříve vybírala autory z oblasti společenských věd. Nejprve jsem si stanovila několik základních kritérií:

1. Autoři musí být z oboru společenských věd a mít zároveň vztah k environmentální problematice;
2. Texty nesmí být zaměřené výlučně na technologie, ale tato problematika by měla být alespoň okrajově přítomna;
3. Tvorba autora by se měla zabývat tématem rizika.

Po zahrnutí těchto kritérií jsem z důvodu vymezení užšího vzorku autorů omezila výběr na ty, které tvořili převážně po druhé světové válce, jelikož první počítač ENIAC, který pracoval podobně jako první počítače byl uveden do provozu v roce 1944 na univerzitě v Pensylvánii. [Wooldridge, 2021: 10] Přidala jsem tedy ještě následující kritérium:

4. Autoři, kteří tvořili hlavně v druhé polovině 20. století

Po uplatnění těchto kritérií jsem z výsledného vzorku vybrala následující autory:



Ulricha Becka pro jeho nadčasovou teorii o rizikové společnosti, Anthonyho Giddense, který se holisticky díval na moderní společnost a konečně Niklase Luhmanna, který ve svém raném díle přímo řeší riziko, ale zároveň čerpal své myšlenky z mnoha oborů, jakým byla například i informatika a kybernetika. Zároveň jsou všichni vybraní autoři relevantní i pro obor sociální a kulturní ekologie, a to je téměř zárukou multioborového pohledu na danou problematiku.

Výběr literatury zabývající se umělou inteligencí nebyl tak přímočarý. Nejprve jsem se snažila vybrat díla, která budou aktuální a nejlivnější v tomto oboru. Z tohoto výběru vzešlo několik nejlivnějších autorů v rámci populárně naučné literatury zabývající se umělou inteligencí. Těmi jsou například Nick Bostrom, Toby Ord a další. Zároveň se ale všichni zabývají umělou inteligencí z pozice určitého diskurzu. Proto jsem se rozhodla tento výběr rozšířit i o počítačové vědce jako například Michaela Wooldridge nebo Stuart Russel a nechala seznam literatury organicky růst tak, jak se objevovala nová témata. Z tohoto důvodu se v této části objevují díla z online pramenů nebo jiných alternativních zdrojů, a to v míře větší, než je v práci tohoto typu zvykem. Jak říká rčení připisované Karlu Čapkovi: „Není nic staršího než včerejší noviny.“ V případě umělé inteligence a jejího rapidního pokroku často ani ty „dnešní noviny“ nejsou schopny držet krok s vývojovými trendy a publikovat tak rychle, aby zůstaly relevantní. Také díky povaze tématu a okruhu čtenářů je online forma nejrelevantnějším médiem.

### 3 Stručný úvod k analýze

Před samotnou analýzou vybraných děl je nutné přiblížit několik pojmů, které s umělou inteligencí souvisí a jejich pochopení nám poskytne hlubší vhled do samotného problematiky výzkumu. Těmito pojmy budou: riziko, které se objevuje v debatách o AI, důvěra, jelikož její přítomnost nebo absence ovlivňuje postavení umělé inteligence ve společnosti a konečně krátce i predikce budoucnosti, jelikož o její naplnění nebo nenaplnění v dílech společenskovedních autorů jde. Podklady pro tyto pojmy jsem proto čerpala hlavně z díla vybraných autorů.

#### 3.1 Riziko

Umělá inteligence již dávno není jen abstraktním pojmem ze sci-fi filmů, ale během velmi krátké doby se stala i nedílnou součástí našich životů. Zvláště pak v posledních letech se setkáváme se stále významnějšími pokroky v této oblasti a přísliby ještě větších benefitů. Oblasti, o kterých se v této souvislosti mluví jsou opravdu rozličné a sahají například od debat o autonomních autech přes použití v diagnostické medicíně až po řešení změn klimatu. S rychlým postupem ve vývoji a stále komplexnějšímu fungování, ale přichází i reálné obavy z rizik s AI spojených.

Mohlo by zdát, že rizikem se lidé zabývali odjakživa, jelikož představa dob minulých je často spojená s mnohými nástrahami, nebezpečím, ale riziko jako pojem má kratší historii. Představa rizika jako takového se do povědomí společnosti dostala v šestnáctém a sedmáctém století v souvislosti se zaoceánskými plavbami. V této době souvisela s rozvojem pojistného systému v souvislosti s potřebou „ošetření“ rizik plynoucích z lodní dopravy. Samotné slovo riziko<sup>19</sup> se zřejmě do angličtiny dostalo díky španělštině a portugalštině, kde se používalo v souvislosti s plavbou v neznámých vodách. Bylo tak nejdříve spojeno s prostorem a až později se přidala dimenze času, kdy se o riziku hovořilo v souvislosti s ochotou přijímat riziko v čase. Orientace na budoucí rozměr rizika měla spojitost s rozvojem kapitalismu, kdy jsou neustále přepočítávány budoucí zisky investic nebo potenciální ztráty. Poté se již používalo k popisu rozličných nejistých situací. [Giddens, 2000: 35]

Právě pojem nejistoty je podle Giddense neoddělitelný od pojmu rizika. Riziko totiž není stejné jako hazard, kdy je vše ponecháno náhodě. Riziko je spojeno s aktivním

---

<sup>19</sup> V angličtině jako “risk”

posuzováním dané situace vzhledem k možným scénářům a budoucím možnostem. Je proto typické pro společnost, která se zaměřuje na budoucnost a snaží se jí kontrolovat. [Giddens, 2000: 36] Ve vztahu rizika a budoucnosti se shodují i Beck [Beck 2004] a Luhmann [Luhmann, 1993] V dílech obou autorů je riziko popisem budoucnosti, které ale zároveň ovlivňuje přítomnost a fungování společnosti.

Zatímco Giddens zmiňuje rozdílnost rizika a hazardu, Luhmann důrazně odděluje mezi rizikem a nebezpečím<sup>20</sup>. Podle něj toto rozdělení nevychází z lidského rozhodnutí, typu ztráty nebo pravděpodobnosti, ale podle původu negativní události. Riziko, je podle něj odlišné od nebezpečí tím, že na rozdíl od nebezpečí nejsou případné budoucí ztráty přičítány vnějším faktorům, ale jsou vnímány jako důsledek určitých rozhodnutí. [Luhmann, 1993: 101]

Luhmann obecně bere jako zdroj narůstajících rizik a ohrožení právě nutkání moderní společnosti dělat stále další rozhodnutí tváří v tvář nejisté budoucnosti. [Luhmann, 1993] Říká totiž, že rizika se vynořují pouze jako komponent rozhodnutí a činů, a tedy samy o sobě neexistují. A přichází s tvrzením, že riziko nenastane, pokud se zdržíme rozhodnutí. Toto tvrzení mírní tím, že samozřejmě jsou situace, kdy se může vyplatit rozhodnutí učinit nebo kdy dokonce nic jiného nezbyvá, ale stále je to pouze o naší<sup>21</sup> volbě. [Luhmann, 1988: 100] S tímto tvrzením nesouhlasí Giddens [2003: 36] a říká: „Nečinnost je často riskantní a existují rizika, kterým musíme čelit všichni, ať chceme, nebo nechceme [...]“ Podle něj, ať se na problém díváme z jakékoli strany, vždy to vede k nutnosti zvládat rizika. Apeluje tak na vlády, aby s rozšiřujícím se rizikem přestaly předstírat, že je to mimo oblast jejich kompetencí. Zmiňuje také důležitost spolupráce i na mezinárodní a globální úrovni, jelikož novodobá rizika neznají hranic států. Zároveň nezapomíná ani na roli jedinců, kdy říká, že nemůžeme čekat, až se objeví vědecké důkazy. [Giddens, 2000: 48]

Část Luhmannova tvrzení o „naší volbě“ jde i proti tomu, jak moderní rizika vnímá Ulrich Beck. Ten říká, že narozdíl od bohatství, kdy je vždy vyžadována naše volba spojená s jednáním, jsou rizika všude implicitně a nebrzděny svobodným rozhodnutím. O nich mluví z důvodu jejich připsání jedinci jako o *askriptivních rizicích*, kdy ve vyspělé civilizaci existuje jakýsi osud ohroženosti, do něž se člověk rodí, nemůže mu uniknout a týká se všech bez výjimky. Podle Becka pak tváří v tvář takovým rizikům může individuální možnost

---

<sup>20</sup> V angličtině jako risk and threats

<sup>21</sup> Luhmann v angličtině používá výraz “one’s own choice”, což implikuje individuální možnost rozhodnutí

rozhodování jen sotva obstat, jelikož zdroje ohrožení jsou v industriálním světě pevně začleněny i do elementárních životních procesů. [Beck, 2004: 53–54] Beck mluví i změně povahy rizik v čase. V minulosti byla rizika vázaná na jedince a mohla tak být spojena s dobrodružstvím a odvahou, moderních rizika mají povahu globálních ohrožení s možností zničení života na zemi. [Beck, 2004: 28] Zatímco v dřívějších etapách bylo možné situace ohrožení připisat vnějším příčinám, v dnešní společnosti to již nelze. [Beck, 2004: 301] Rizika modernizace se projevují současně jako lokálně specifická a jako nespecificky univerzální a ukazuje se, jak nepředvídatelné a spletité jsou jejich cesty. Podle Becka rozdělení rizik, stejně jako rozdělení bohatství, kopíruje třídní schéma, jen obráceně. Riziko se totiž narozdíl od bohatství shromažďuje dole. Zvláště pak u „klasických“<sup>22</sup> typů rizik, rizika třídní společnost upevňují, jelikož nedostatek zdrojů je spojen s nedostatkem bezpečí. Pro tento typ rizik stále platí, že je možné se odstěhovat dále od průmyslové zóny, pořídit si filtry vzduchu nebo kupovat kvalitnější potraviny. Toto třídní rozdělení však narušují modernizační rizika. U nich postupně kvůli globalizaci mizí cesty, kudy se lze z rizika pomocí peněz „vykoupit“. Taková rizika mažou společenské třídy a staví všechny před stejné ohrožení. Jsou v tomto ohledu demokratizační, nemají hranice [Beck, 2018: 48]

Beck říká [Beck, 2018: 36]: „U modernizačních rizik je to, co je od sebe obsahově a věcně, prostorově a časově odděleno, uváděno do kauzálního vztahu a zároveň i do kontextu sociální a právní odpovědnosti.“ Tyto předpokládané kauzality však nadále zůstávají teoriemi a jako takové musí být pokládány za pravdivé a musí se v ně věřit. [ibid.] Společnost je konfrontována sama se sebou a riziky, která jsou jejím vlastním produktem. Zdrojem ohrožení tak není nevědění, jako to bývalo v minulých epochách, ale naopak vědění, které pokračuje v tradici spojené s ovládnutím neovládnutého a překonáním nepřekonaného. Stává se tak současně hrozbou a zároveň příslibem k jejímu zvládnutí. [Beck, 2004: 301]

Riziko je často spojováno s negativními situacemi a stavy ohrožení se kterými se musí člověk v životě vypořádávat a vidí je tak například Beck. Giddens si všímá nejen negativních aspektů rizika, ale i těch pozitivních, kdy je riziko „mobilizující hnací silou společnosti, která si přeje změnu – která chce svoji budoucnost určovat, a ne ji nechávat na náboženství, tradici nebo rozmarech přírody.“ [Giddens, 2000: 36] Giddens v této souvislosti rozlišuje riziko do dvou typů. V případě prvního, *vnějšího rizika* jde o situace, které jsou tradičně součástí lidského života a společnost se s nimi setkává od počátku věků.

---

<sup>22</sup> Myšleno jiných než typických pro moderní dobu

Můžeme si například představit situace spojené s přírodou a jejími jevy, kterým čelili lidé vždy. Od něho je však nutné odlišit druhé, *vyrobené riziko*, které si vytváří moderní civilizace sama. Pro druhý typ je typické, že s nimi máme jen malé historické zkušenosti. Příkladem by mohli být opět přírodní jevy, ale na rozdíl od předchozí situace jde například o globální změny klimatu, tedy jev, za které si společnost přivodila sama, často i snahou o kontrolu přirozených přírodních projevů. Právě tento přechod od převahy rizik vnějších k rizikům vyrobeným je charakteristický pro současnou společnost. Neplatí to však pro všechny stejně, jelikož v chudších zemích je stále více „tradičních“ rizik, kdy je více zasahují například rizika plynoucí z nedostatečné úrodu nebo vodního zdroje, ale i to se vlivem globalizace mění [Giddens, 2000: 40]

Během posledních několika desetiletí se oblast řízení rizik začlenila do procesů fungování v různých odvětvích, jelikož každá organizace se s nějakým rizikem musí potýkat. Tato rizika se netýkají pouze přímého působení činnosti například v průmyslu, ale i toho, jak vnímají aktéři rizika spojená s organizací a jak následně jednájí. Proto se čím dál více rozvíjí oblast PR a rizikové komunikace, která se snaží sociální aspekt rizik zvládnout a předejít situacím, kdy by veřejné mínění mohlo samo o sobě produkovat riziko.

Riziko už tedy není spojeno jen se specifickými odvětvími jako například nukleární technologie nebo genetické inženýrství. Stalo se oblastí, se kterou se musí zabývat nejen všechny soukromé, státní i neziskové organizace, ale týká se čím dál více i jednotlivců. Ti v individualizované době bez jistot, tradic a hodnot, které měli předchozí generace, musí být aktivním tvůrcem svého života a rozhodovat se o čím dál více komplexnějších problémech v podmínkách vznikajících stále dále od nich. [Beck; Beck–Gernsheim, 2002: 23] [Giddens, 2003] Tímto posunem ve vnímání rizik se jím, jakožto zdrojem ohrožením a zároveň hybatelem moderní společnosti, se začala pochopitelně zabývat sociologie.

## 3.2 Důvěra

Důvěra je silně spojena s pojmem rizik, kdy funguje jako jeden z mechanismů vyrovnání se s představou rizika. Velký sociologický slovník ji definuje jako: „[...] typ postoje a zároveň mezilidského vztahu, který vyvolává pocit jistoty plynoucí z přesvědčení, že partner komunikace (osoba, instituce) splní určitá očekávání.“ [Maříková, Petrušek, Vodáková 1996: 231] První, kdo se zabýval tímto pojmem jako samostatnou oblastí byl německý filosof Georg Simmel. Ten ji vnímal jako zásadní součást sociálních vztahů.

S tématem důvěry se nedomyšlitelně pojí také jméno Niklase Luhmanna. Ten přišel s rozdělením pojmů na *důvěru*<sup>23</sup>, *důvěřivost*<sup>24</sup> a *obeznámenost*<sup>25</sup>. Nejprve rozlišuje mezi obeznámeností, která je nevyhnutelným faktem a důvěrou, která je v jeho očích řešením pro specifický problém rizika. Obeznámenost nebo také známost se učíme od prvního momentu, kdy se narodíme a naše prostředí začneme rozlišovat na známé a neznámé. S tím, jak postupně procházíme světem a děláme další taková dělení, utvoříme si kolem sebe svět známého, který je ohraničen neznámým. Můžeme žít ve známém světě i díky tomu, že dokážeme znovu začleňovat neznámé do světa známého. Zároveň však díky známým vztahům, postupům a vědění stále víme, co je pro nás neznámé. V tom nám pomáhají symboly, které funkčně nahradili mýty, které měli stejné poslání. Symboly tradičně byly oblastí náboženství. V moderní době ale přibyl nový pojem – riziko. To dalo společnosti najevo, že neočekávané události mohou být výsledkem našich rozhodnutí, a ne pouze zásahem vyšší moci, přírody nebo Boha v závislosti na kosmologii dané společnosti. Luhmann ale zmiňuje, že nenahradilo funkci náboženství, jen se přidružilo jako další aspekt života. Důvěra a důvěřivost jsou pak umístěny symbolickými reprezentacemi do známého světa a mohou být zničeny nebo ohroženy v případě událostí otrásajících danou symbolikou. [Luhmann, 1988: 95–97]

Zásadní je ale podle Luhmanna rozdělení důvěry a důvěřivosti. V obou případech jde o očekávání, která mohou skončit zklamáním, ale přeci se podle něj liší. Běžný je stav důvěřivosti, kdy jsme přesvědčení, že naše očekávání nebude zklamáno. Věříme, že se politici budou snažit zabránit konfliktům, že nám v ruce nevybuchne telefon. Jsou zkrátka situace, kdy nezbyvá nic jiného než věřit, a tak povětšinou činíme. Je to způsobeno jak tím, že pravděpodobnost takové události je malá, ale také díky tomu, že nám nic jiného nezbyvá. Luhmann říká, že jedinou alternativou by bylo žít v neustálé nejistotě a nemít žádná očekávání.<sup>26</sup> [Luhmann, 1988: 97] Důvěřivost má tedy spíše pasivní povahu. Na druhou stranu důvěra, vyžaduje předchozí účast a lze o ní mluvit pouze v případě, že možné nežádoucí následky mohou být větší než benefity. Znamenala by například vědomé koupení nebo nekoupení levné nabíječky telefonu, cestování nebo necestování stopem. V takovém

---

<sup>23</sup> V originále: “trust”,

<sup>24</sup> V originále: “confidence”,

<sup>25</sup> V originále: “familiarity”,

<sup>26</sup> Je otázkou, jestli, popřípadě jak by tímto pohledem otrásl poslední dva roky, kdy se hned několik neočekávaných událostí stalo realitou.

případě se podle Luhmanna člověk riziku vyhne, ale zároveň se vzdává výhod, která by z riskantní situace mohla vzejít jako by v uvedeném případě byla úspora peněz nebo nové zážitky. Rozdíl mezi oběma probíranými pojmy je tedy ve vlastním vnímání situace. Pokud tedy nezvažujeme alternativy (dražší nabíječku, cestu hromadnou dopravou), pak jde o situaci *důvěřivosti*. Pokud se rozhodneme nějakou volbu preferovat navzdory možnosti, že budeme zklamáni, jde o *důvěru*. A pokud zklamání nastane, pak v případě důvěřivosti budeme zklamáni externími faktory, případě důvěry těmi interními a případně litovat naši volby. [Luhmann, 1988: 97–98]

Důvěrou se zabýval také Anthony Giddens, který reaguje na předchozí Luhmannovi teze. Souhlasí s rozdělením důvěry a důvěryhodnosti, ale není spokojen s detaily Luhmannova pojetí důvěry výše. Například premisa, že v rámci důvěry jedinci zvažují konkrétní alternativní jednání je podle něj nesprávná, jelikož důvěra podle Giddense kontinuálně vyplývá z předešlých situací. Také mezi důvěřivostí a nebezpečím podle něj neexistuje vnitřní spojení. Nebezpečí je přítomno za podmínek rizika a pro jeho definici má význam. Giddens uvádí příklad, kdy riziko při plavbě Atlantikem bude podstatně větší, pokud budeme v malém člunu, než kdybychom byli na zaoceánské lodi, a právě to je dáno odlišnou mírou nebezpečí. V reakci na předchozí přichází Giddens s deseti body definujícími důvěru.

1. Důvěra je vztažena k nepřítomnosti času a prostoru. Tedy není potřeba důvěry v naprosto transparentní situaci.
2. Důvěra není spojena s rizikem, ale s nahodilostí. Navzdory nahodilým výsledkům je důvěra spojená se spolehlivostí, a to jak jedince, tak i organizací.
3. Není ale totéž, co víra ve spolehlivost jedince nebo systému, ale tím, co od této víry odvozujeme.
4. Je možné hovořit o důvěře v symbolické znaky nebo expertní systémy<sup>27</sup>, ale jde o důvěru ve správnost principů ne v morální bezúhonnost ostatních.
5. Důvěra může být definována jako důvěřivost (confidence) ve spolehlivost osoby nebo systému týkající se určitého souboru výsledků nebo událostí. Důvěřivost vyjadřuje víru v přesnost abstraktních principů u technologií.

---

<sup>27</sup> Viz kapitoly níže

6. V podmínkách modernity důvěra existuje v kontextu obecného povědomí, že lidská činnost je spíše vytvářena sociálně, než dána povahou věcí a božským vlivem. Pojem rizika nahrazuje pojem štěstěny, ale znamená to spíše změnu ve vnímání určenosti a nahodilosti v tom, že místo náboženských kosmologií působí lidské morální imperativy, přírodní příčiny a náhoda.
7. Nebezpečí a riziko jsou úzce spojené, ale nejsou totožné. Rozdíl není v tom, jestli jedinec při určitém jednání zvažuje nebo nezvažuje alternativy. Riziko předpokládá přesné určení nebezpečí, ale ne nutně jeho vědomí. Je možné tedy jednat v riskantních situacích, ale riziko si neuvědomovat. V takové situaci si nejsou vědomi nebezpečí, které jim hrozí.
8. Riziko a důvěra se navzájem proplétají. Důvěra obvykle slouží ke snížení nebezpečí a v určitých okolnostech jsou struktury rizika institucionalizovány. S rizikem se také kalkuluje a to, co se jeví jako ještě přijatelné riziko se liší v závislosti na kontextu.
9. Riziko není jen individuální záležitostí. Rizika jsou i taková, která hromadně ohrožují masy. Stav bezpečí je tedy situace, ve kterém se působí proti specifickému souhrnu nebezpečí a v níž je minimalizováno. Zkušenost bezpečí závisí na rovnováze důvěry a přijatelného rizika.
10. Předcházející nevypovídá nic o opaku důvěry, kterým není nedůvěra. Také nic neříkají o tom, a jakých podmínek se důvěra vytváří.

1. [Giddens, 2003: 36–39]

Pokud oba přístupy shrneme pak vyplývá, že důvěra je spojena s očekáváním nějaké situace, výsledku, chování. Luhmann ji spojuje s konkrétním rozhodnutím vázaným na daný okamžik. Pro Giddense má důvěra mnohem kontinuálnější povahu. Není vázána jen na jedince a v době pozdní modernity, kdy dochází k rozpojení času a prostoru je důvěra v abstraktní systémy jako je například i expertní vědění, kterým se budu zabývat níže, nezbytná.

### 3.3 Budoucnost

Umělá inteligence je i díky reprezentaci v kultuře vnímána jako technologie budoucnosti. Do budoucnosti jsou zasazeny i obavy z jejího nekontrolovaného působení, které by mohlo ohrozit samotnou existenci lidstva.



Ve všech úvahách nad rizikem je neodmyslitelně zahrnuta i budoucnost. Právě v ní se odehrávají situace, kterým se člověk nebo společnost snaží předejít nebo zmírnit jejich následky. Díky tomu mají rizika, ač v důsledcích velice reálná, v sobě i něco nereálného. To zmiňuje Beck, když říká: „V rizikové společnosti ztrácí minulost svoji determinující moc nad přítomností. Na její místo nastupuje jako „příčina“ přítomných prožitků a dnešního jednání budoucnost, a tudíž něco neexistujícího, vykonstruovaného, fiktivního.“ [Beck, 2018: 44] Podle Luhmanna moderní společnost prožívá svou budoucnost v rámci rizik plynoucích z dnešních vlastních rozhodnutí. „Riziko je aspektem rozhodnutí a rozhodnutí mohou být učiněna pouze v přítomnosti.“<sup>28</sup> [Luhmann, 1998: 71]

Giddens tuto myšlenku rozšiřuje o poznatek, že riziko přímo předpokládá společnost orientovanou na budoucnost. Vyvrací častou domněnku, že riziko není specifickým rysem naší doby, jelikož nebezpečí tu bylo vždy. Tato představa je podle něj často spojena s vizí života v minulosti plného všudypřítomného nebezpečí, krutosti a neblahých vyhlídek na jeho dlouhé trvání. [Giddens, 2000: 34–35] To však zpochybňuje a nastiňuje vývoj vnímání nejistých situací až po riziko, jak ho známe dnes. Tradiční společnosti podle něj pojem rizika neměly, neb ho nepotřebovaly. Pokud se něco špatného stalo, braly to jako něco, co se stává nebo jako vůli bohů. Je tedy nezbytné jasně oddělit nebezpečí nebo hazard od pojmu rizika, jelikož to není totéž. „Riziko se vztahuje k nejistým situacím, které se aktivně posuzují vzhledem k budoucím možnostem. K širokému užívání slova dochází jenom ve společnosti, která se orientuje na budoucnost – která se na budoucnost dívá jako na území, jež je třeba dobýt nebo kolonizovat.“ Touhu po odtržení se od minulosti dokonce vnímá jako základní charakteristiku moderní průmyslové civilizace. [Giddens, 2000: 35–36]

Zatímco v minulosti bylo ohrožení a riziko bráno jako zásah vyšší moci a společnosti se s nimi vypořádávaly skrze rituály, magii a víru, v moderní společnosti se rozvíjelo pojišťovnictví a výpočet pravděpodobností. O pojištění píše Giddens, ale zmiňuje ho ne tak ve smyslu soukromého nebo komerčního pojištění, jako spíše důležitého činitele díky kterému jsou lidé ochotni nést budoucí riziko. To je možné právě díky moderní společnosti, která věří ve svou kontrolu nad budoucností. Právě víra je i pro samotnou existenci pojištění důležitá, jelikož se sjednáním pojištění reálné riziko nezmizí, jen se jeho následky převedou na někoho jiného. Jde tak pouze o redistribuci rizika, a to nejen v rámci jednotlivců a firem, ale i sociálního státu nebo napříč kontinenty. [Giddens, 2000: 38–39] S tím souhlasí i

---

<sup>28</sup> V originále: “A risk is an aspect of decisions, and decisions can be made only in the present.”

Luhmann, který zmiňuje, že pojištění má jen pramálo co dočinění se zvládním rizik, jako spíš se zárukou, že nežádoucí událost neohrozí oběť finančně. [Luhmann, 1998: 72] Pojištění nás tedy před rizikem neochrání, ale co samotné varování před rizikem?

Luhmann upozorňuje na paradox varování, kdy pokud jsou vyslyšena, společnost se podle nich zařídí a obávaná situace nenastane. Vystávají otázky, jestli by vůbec nastala, pokud bychom varování nevyslyšeli. [Luhmann, 1998: 76] S podobnou myšlenkou přichází I Giddens, když říká, že ačkoli panikaření často vede k tomu, že díky přehodnocení svých rozhodnutí situace nakonec nenastane, může právě úspěšné odvrácení událostí být důvodem pro pochybnosti, jestli by ohrožení vůbec nastalo. Dodává, že ve většině případů vyrobeného rizika se bude zpochybňovat i to, jestli vůbec o riziko jde. Souvisí to podle něj s faktem, že dopředu nikdy nevíme, jestli je panika zbytečná nebo není. [Giddens, 2000: 45]

S vizí budoucnosti se tedy dnes snažíme aktivně předejít nebo zmírnit dopady našich rozhodnutí a činů současných i minulých. Luhmann říká, že budoucí přítomnosti závisí na rozhodnutích, která musíme udělat nyní. Dodává, že rozhodnutí můžeme činit pouze pokud to, co nastane je nejisté. [Luhmann, 1998: 67] Jenže jak vlastně předvídáme budoucnost? Podle Luhmanna představy o budoucnosti závisí na společnosti, která je formuje. Nejsou tedy vynálezem moderní společnosti a její předvídaní se datuje až do období Mezopotámie. Až do počátku moderní doby byla budoucnost vnímána jako dokonalá souhra konstant a variací nahodilých událostí. Cokoliv, co člověk podnikl, mohlo dopadnout dobře nebo špatně a mohl dokonce zemřít, ale nijak to neovlivnilo existenci lidstva. Nejistoty budoucnosti zůstávaly stále v rámci fungování světa jako souhry viditelných a neviditelných věcí. Tento model však nemohl být zachován v moderní společnosti, kde s narůstajícím věděním narůstala i její komplexita. V této době se také posunula od představy světa jako dokonalého o představám k jeho zdokonalování. [Luhmann, 1998: 63–65] V té samé době se objevuje víra v budoucnost a neustálý pokrok. Tato tradice má původ v osvícenství, které se postavilo náboženskému dogmatu. Osvícenští myslitelé razili zásadu, že čím více poznáme svět a osvobodíme se od zvyků a tradic minulosti, tím více budeme mít moc nad utvářením přítomnosti a budoucnosti. Podobné tvrdil o mnoho let později Karel Marx. V duchu této představy měl svět s dalším rozvojem vědy a technologií nabývat čím dál větší stability. Giddens v tomto zmiňuje spisovatele George Orwella, který od budoucnosti očekával společnost tak stabilní a předvídatelnou, že v ní člověk bude mít místo pouze jako kolečko skvěle namazaného stroje ekonomiky. [Giddens, 2000: 14] Jak však sami víme z vlastní zkušenosti, dnešní svět se těmto představám příliš nepodobá. Vědění a technologie, které nás měli osvobodit od minulosti a předurčenosti, měli na společnost spíše opačný

účinek. Přinesly mnohem více komplexity, nad kterou máme jenom pramálo kontroly, jak si všímá Luhmann, rizika, před kterými není uniku, jak popsal Ulrich Beck a odpoutání od minulosti nahradila transformace času a prostoru, které se věnuje Anthony Giddens.

## 4 Analytická část

V následující části práce se budu věnovat analýze samotných děl. Jak bylo zmíněno výše, budu se zabývat především vybranými díly Ulricha Becka, Anthonyho Giddense a Niklase Luhmanna. V analýze není obsaženo celé jejich dílo ani teorie, jelikož to by významně překročilo rozsah diplomové práce. Budu se tedy věnovat pojmům a oblastem, které mi po prostudování literatury přišly nejnositelnější.

Výběr oblastí, kterým se věnuji podrobněji není nahodilý, ačkoli by se takový mohl zdát. Jelikož předmětem mého zkoumání byla díla třech společenskovedních velikánů zabývajících se i environmentální tematikou, která nutně spojuje holistický pohled na svět. Tento výběr dává dle mého soudu ve výsledku ucelený obrázek společnosti, v jaké dnes žijeme a jakou umělá inteligence přetváří.

Prvním tématem je věda a její utváření. Jak bylo popsáno v teoretické části, i díky postavení oboru umělé inteligence a vysokým vstupním nárokům na expertní znalosti, už od počátku utváření této technologie vznikala bariéra mezi vědeckým světem a tím veřejným.

### 4.1 Věda

#### 4.1.1 Věda v rizikové společnosti

Věda zajisté zajistila společnosti růst pokroku a blahobytu, ale s jejím rozvojem přišla i řada problémů. Prvním člověkem, který si dal do spojitosti vědu a zlepšení životních podmínek byl v 16. století Francis Bacon. Ten dokázal přiblížit vědu většímu množství lidí a přesvědčovat je o benefitech pokroku a poznání. [Postman, 1993: 35] Tématu vědy jako vědní disciplíny se zabýval například Thomas Kuhn, který v jeho knize *Struktura vědeckých revolucí* popisoval vědu spíše než jako lineární proces, dynamické střídání paradigmat. Také rozpracovává téma rozdílu přírodních a společenských věd, které bylo předmětem mnohých akademických sporů jeho doby. [Kuhn, 1997] Jenže věda nebyla jen o poznání světa, ale i aktivně měnila svět. S obdobím společenských změn 60.let přišla i vlna kritiky vědy. Ta upozorňovala na objektivitu vědy, téma, kterým se zabýval například francouzský sociolog a filosof vědy Bruno Latour a mezi jinými i právě Ulrich Beck, z jehož *Rizikové společnosti* budu nejvíce čerpat pro tuto pasáž analýzy.

Jak již bylo řečeno v kapitole 4.2 věnující se riziku, s přechodem od tradiční společnosti přibývá i rizik, ale také se mění jejich povaha. Zatímco dříve měla charakter osobního ohrožení, dnes mají globální dosah a ohrožují všechny bez rozdílu. [Beck, 2004:

28] Také se zcela proměnil původ rizik. Dříve bylo nebezpečí a riziko podmíněné externě (bohy, přírodou), to se ale razantně proměnilo s nástupem moderní společnosti, kdy rizika začala být vědecky a sociálně konstruovaná<sup>29</sup>. Beck dává vztah rizika a vědy do následujících trojích souvislostí [Beck, 2004: 257]: „věda se stává (spolu)příčinou rizik, prostředkem pro jejich definování a zdrojem pro jejich řešení, a právě tím si otevírá nové trhy *scientizace*.“ Tu rozděluje na *prostou*, kdy aplikujeme vědu na svět přírody, který je daný a *reflexivní*, kdy je věda konfrontována i sama se sebou. Během prosté scientizace je konfrontována tradice se světem moderny a laikové a experti. V období, které trvá až do první poloviny dvacátého století, je přítomna až neotřesitelná víra ve vědu a pokrok. Zároveň díky svým objevům má věda moc umlčet případné skeptiky a kritiky. [ibid.] To se však mění s příchodem reflexivity ve vědě, kdy musí čelit sama sobě i rizikům, které vyprodukovala, ačkoli nás před nimi měla chránit, a slibům, které nedokázala naplnit. Kritika přichází nejen z oblasti vědy, například té sociální jako v případě Kuhna, ale i zprostředkovaná veřejnou sférou. To otřásá do té doby pevnými základy vědecké sféry a odhaluje své nejistoty. Tím se uvádí do chodu proces demystifikace věd<sup>30</sup>. [Beck, 2004: 259] Díky němu dochází k demonopolizaci nároků věd na poznání, kdy se věda stává stále nutnější jako v případě neviditelných rizik, o kterých budu mluvit níže, ale také již není společensky závaznou pro definici pravdy.

Beck se ve svém díle zabývá i postavením laika v takovém systému vědy, kdy v rámci prosté scientizace byl laik vytlačen mimo oblast vědy a měl se stát pouze jejím příjemce nebo objektem. [Beck, 2004: 261] Giddens uvádí příklad nemoci šílených krav ve Velké Británii, kdy na možné nebezpečí upozorňovala různá hnutí a veřejnost, ale nebyli uposlechnuti. Říká, že se epidemii mohlo zabránit, pokud by se zahájil veřejný dialog o technologické změně a jejích možných negativních následcích. Konkrétně zmiňuje, že by veřejnost měla mít více kontaktu s vědou a technologií, jelikož by to pravděpodobně nezabránilo všem rizikům, ale alespoň část by jich zmírnilo. [Giddens, 200: 49-50] Zmiňuje také rizika plynoucí z panikaření a utajování rizik. Právě v přístupnosti informací může AI pomoci, ale v konečném důsledku i uškodit.

---

<sup>29</sup> Se stejným dělením přichází i Giddens, který používá termíny *vnější riziko* a *vyrobené riziko*. [Giddens, 2000: 40]

<sup>30</sup> S tímto procesem přichází i řada děl jako například antropologická studie *Laboratory Life* z roku 1979 Bruna Latoura nebo jeho *Nikdyjsme nebyli moderní* z roku 1991.

Pokud bychom brali umělou inteligenci pouze z hlediska vědy jako oboru informatiky, můžeme najít dva postoje. Na jedné straně měla AI od počátku až do velice nedávné doby vysokou vstupní překážku v podobě vysoce expertního vědění sahajícího od oblastí informatiky, přes pokročilou statistiku až po matematiku. To zcela přirozeně vytvářelo jakousi fyzickou bariéru v průniku laiků a veřejnosti obecně do tohoto oboru. Myslím, že i díky tomu bylo možné, aby se kolem umělé inteligence, u které už i název budí emoce a fantazie, rozvíjela představa AI jako něčeho tajemného, nedotknutelného a mimo naše chápání. Právě díky tomu mohla vzniknout podoba AI dále reprezentovaná v kultuře, a zvláště pak žánru sci-fi, která dle mého názoru významně formovala představy veřejnosti o jejich schopnostech a rizicích AI. To trvalo až do nedávné doby, kdy se díky *opensource*<sup>31</sup> komunitě dostali i k laikům bez odborných znalostí různá dostupná uživatelská rozhraní umožňující přístup k modelům využívajícím umělou inteligenci<sup>32</sup>. Podobné mosty mezi světem umělé inteligence a veřejností dle mého mají potenciál narušit barikády do nedávna pro běžného člověka neuchopitelné a neviditelné technologie AI. To by mohlo přivést i veřejnost do debat o umělé inteligenci a tyto debaty učinit věcnější. A sama umělá inteligence by demystifikací vědy a začleněním veřejnosti mohla pomoci.

Luhmann zmiňuje schopnost moderních technologií zpochybnit autoritu expertů. Dokonce jde dál a přichází s následující predikcí: „V zásadě každý si bude v budoucnu moci zkontrolovat výroky expertů, jako jsou například lékaři a právníci, na svém počítači.“<sup>33</sup> [Luhmann, 2012: 187] Je nutné zmínit, že s tímto trefným tvrzením přichází v roce 1997, kdy *Die Gesellschaft der Gesellschaft* poprvé vychází. Pro kontext, v roce 1997 neexistovaly globální vyhledávače odkazující na celý internet, jako je dnes například Google nebo Yahoo. Uživatel musel hledat zvláště informace, soubory nebo texty a zadat konkrétní server. [Procházka, 2011] Je tedy jasné, že za těchto podmínek nebyl přístup k informacím tak jednoduchý a samozřejmý, jako je tomu dnes. O přínosu *rozumové emancipace* mluví Ulrich Beck, ten říká, že lidé se mohou vysvobodit z kognitivního diktátu expertů. Za to podle něj může fakt, že věda už není nedotknutelná a posvátná, ale může být kritizována. Beck [Beck,

---

<sup>31</sup> Opensource se používá pro zdrojový software s otevřeným zdrojovým kódem. Kolem této oblasti se vytvořila komunita prosazující princip otevřenosti do různých oblastí informatiky.

<sup>32</sup> Příkladem je program AI DALL·E 2 vytvořený společností OpenAI, který umožňuje vytvářet obrázky z textového zadání. Jeho tvorbu je možné nalézt v příloze této práce.

<sup>33</sup> V originále: „In principle, everyone will in future be able to check the statements of experts such as physicians and lawyers on his own computer.“

2018: 278] dodává: „Rozdíl mezi „neosvícenou lůzou“ a „osvícenými občany“ nebo – moderněji řečeno – mezi laiky a experty mizí a mění se v konkurenci různých expertů.“ Právě v tom napomohla i umělá inteligence, která je schopná díky pokročilým algoritmům zkvalitnit vyhledávání a zlepšit tak přístup k vědomostem. Pokud tak člověk dostane vyjádření nějakého experta, které by si rád ověřil, je to možné doslova během pár kliknutí a není jen odkázán na důvěru v jeho vědění. Také díky stále se zlepšujícím překladačům se boří jazyková bariéra, kdy se najednou člověku otevrou vědomosti, ke kterým by dříve neměl přístup. Pokud bychom šli ještě dále, pak existují aplikace jako například *Google Lens*, kdy je možné vyhledávat pouhým vyfocením hledaného objektu. Všechny tyto technologie tak otevírají přístup ke zdroji informací, který jsme si do nedávné doby nedokázali ani představit. Luhmann v tomto kontextu říká, že počítačem zprostředkované vědění mělo zásadní vliv na společnost, kdy se stala opravdu světovou společností. V tomto kontextu zmiňuje propojení světa sítěmi sbírajícími a zpřístupňujícími data. Ani tato schopnost AI poskytovat nesčetné množství informací, není bez rizika.

V moderní společnosti jsme čím dál závislejší na vědeckých poznatcích, zároveň jsme i díky snazšímu přístupu k informacím nezávislejší na konkrétních vědeckých výsledcích. Díky množství vědeckých studií je však patrná i jejich rozpornost. To umožňuje formování nejen vědy, ale i *antivědy*. Ta nevzniká z nedostatku vědění, ale naopak z jeho přebytku. Taková věda si pak podle Becka redukuje své nároky na poznání do pouhých hypotéz, nabízí *seberelativizující pluralismus interpretací* a je pro ni důležitá *sociální akceptance*. Díky ní již nestačí pouze předkládat výsledky jako holá fakta, ale je nutné prezentovat výsledky sám a vzbudit dobrý dojem. Tím se do oblasti vědy, která byla dříve brána jako objektivní, dostává i sociální aspekt. S tím, jak nutnější je vědecké poznání, ale čím dál méně dostačující. Vzniká tak prostor pro různé oblasti jiného přístupu k vědění, které jsou ale účelově promíšené s vědeckými poznatky, kritikou vědy a zároveň vírou ve vědu. [Beck, 2018: 278] To se projevilo i během posledních let, kdy během koronavirové pandemie se na právně vyděšenou společnost prahnoucí po informacích valilo množství různých studií, poznatků a predikcí, které si často protirečily. I díky tomu se veřejný prostor zaplnit mnoha *novými alchymisty*, jak je Beck nazývá. Nebezpečí takových „odborníků“ je podle něj v tom, že se „dovolávají vědy samé, aby odmítly její nároky.“ [Beck, 2018: 278] Jsou tak díky „ozbrojení vědou“ imunní proti vědě samotné. To se podílelo na rozpolcení společnosti v době, kdy na sobě byla závislejší než kdy jindy. Podobná situace nastala jen o něco později v souvislosti s konfliktem na Ukrajině. Na tom pochopitelně měla podíl

rozpolcenost vědy a globalizace, ale troufám si říci, že i umělá inteligence. Kvůli ní bylo možné na základě algoritmů, a zvláště těch na sociálních sítích, polarizovat názory.

K manipulaci mínění s využitím AI v nejčastějších případech dochází v oblasti sociálních sítí jako je například Facebook. Zpravidla se tak nejvíce děje během voleb či krizových situací, jakými jsou například ozbrojené konflikty. Souvislost s tím mohl předznamenat i Beck

Sociální média fungují na základě algoritmů, která nabízejí obsah každému jednotlivci podle jeho chování na těchto platformách. V konečném důsledku tak vidíme pouze to, co nejspíše vidět chceme. Pokud například uživatel sdílí a interaguje s příspěvkem týkající se cyklistiky, algoritmus si tuto vazbu zapamatuje a bude upřednostňovat příspěvky s touto tematikou oproti té obsahující automobily. To, co funguje u dopravních preferencí, samozřejmě funguje i u politických názorů. Uživatel tak v určitém okamžiku získá dojem, že všichni ostatní si myslí totéž co on, protože nebude vystaven jiným názorům. Toto vede ke vzniku nebezpečných *sociálních bublin*. Uživatelé v nich pak kvůli algoritmu jen zřídka nebo vůbec nevidí protiargumenty nebo příspěvky, které by dané rétorice odporovaly. Díky pokročilým textovým modelům je navíc možné vytvořit velké množství obsahu velmi rychle [Buchanan; Lohn; Musser; Sedova, 2021] a šířit ho po sociálních sítích pomocí internetových botů<sup>34</sup> pracujících 24 hodin 7 dní v týdnu. Jakmile tyto falešné a zavádějící zprávy převezmou a začnou šířit i ostatní uživatelé je velmi obtížné je zastavit. V období tří měsíců před prezidentskými volbami v USA v roce 2016 tak 20 nejčastějších dezinformací na Facebooku vyvolalo více komentářů, sdílení a lajků než 20 nejčtenějších zpráv z 19 hlavních zpravodajských zdrojů dohromady (včetně The Huffington Post, The New York Times a The Washington Post). Tři čtvrtiny těch, kteří tyto falešné zprávy četly, věřily, že jsou pravdivé. [Silverman, 2016] Podobná situace nastala i v České Republice v roce 2018 při prezidentských volbách, kdy se mediálním prostorem šířilo velké množství dezinformací o obou kandidátech. [ČT24, 2016] K ilustraci, v jakém měřítku dochází k manipulaci veřejného mínění může sloužit i výzkum vědců z univerzit v Berkeley a Swansea ohledně referenda o Brexitu. Dle zveřejněných dat bylo k ovlivnění hlasování použito přes 156 tisíc ruských účtů na Twitteru a jen v posledních 2 dnech referenda tyto účty vytvořily přes 45 tisíc "tweetů".

Možná ještě nebezpečnější je však schopnost současných systémů umělé inteligence vytvářet a manipulovat audiovizuální obsah. V roce 2018 výzkumníci ze společnosti Nvidia,

---

<sup>34</sup> Internetový bot (od robot) jsou digitálním nástrojem pro provádění zadaných repetitivních úkolů.



zabývající se výrobou počítačových čipů, představili software využívající umělé inteligence k syntéze zcela realistických portrétů osob, které nikdy neexistovaly. Portréty byly vytvořeny novým typem neuronové sítě – generativní kontradiktorní sít' (GAN). Tato revoluční technologie byla mimo jiné o 2 roky později schopna “vytvořit” i legendární hru Pac-Man bez herního software<sup>35</sup>. [Shen, 2018]

Dle studie jsou navíc takto synteticky generované tváře<sup>36</sup> důvěryhodnější než ty skutečné. To může být způsobeno tím, že syntetické tváře mají tendenci vypadat více jako průměrné tváře, které jsou samy o sobě považovány za důvěryhodnější<sup>37</sup>. Snadný přístup k takto vytvořeným portrétům<sup>38</sup> může mít negativní dopady například pro vytváření falešných profilů na sociálních sítích či seznamkách. V případě podvodných videí a obrázků, tzv. deep fakes<sup>39</sup> jsou programy a aplikace na jejich vytváření bez větších problémů dostupné na internetu, včetně těch co vytváří video v reálném čase pomocí webové kamery. Rozlišit pravost je v tomto případě pro laika složité. Svědčí o tom studie německého Institutu Maxe Plancka pro informatiku, v jejíž rámci byla skupině respondentů promítána skutečná videa a deepfakes Baracka Obamy, Theresy May a Vladimira Putina. Více než 50 % respondentů věřilo, že falešná videa jsou autentická. V případě Putina je za pravá považovalo dokonce 65 % respondentů. Pouze 80 % z nich považovala pravá videa za skutečná. [Kaplan, 2020]

Stejně jako může být umělá inteligence využita k tvorbě falešných zpráv a k výrobě deepfakes, může je také odhalit, a to nejen v rámci většího přístupu k informacím, jak bylo řečeno výše. Díky AI lze například zpětně analyzovat potenciálně zmanipulovaná data a obnovit video, zvuk a obrázky do původního stavu. Možné použití je také ve vyhledání původních zpráv, z nichž byly ty falešné vytvořené. Některé společnosti takové nástroje používají již dnes: Facebook používá nezávislé subjekty k ověřování faktických informací a Microsoft používá umělou inteligenci k vývoji algoritmů pro ověření důvěryhodnosti a odhalování falešných zpráv. [Kaplan, 2020]

---

<sup>35</sup> <https://blogs.nvidia.com/blog/2020/05/22/gamegan-research-pacman-anniversary/>

<sup>36</sup> Viz například stránka <https://thispersondoesnotexist.com>.

<sup>37</sup> <https://www.pnas.org/doi/10.1073/pnas.2120481119>

38

<sup>39</sup> Deepfakes jsou fotky nebo videa, které byly pozměněny, neurální sítí aby zahrnovali osoby, které na originálu nejsou. [Wooldridge, 2021: 157-158]

#### 4.1.2 Pozice laika v expertních systémech

Specifickým postavením experta v rámci vědy a vědění moderní společnosti se zabývá i Anthony Giddens. Giddens mluví o dvou *vyvazujících mechanismech*, které umožňují vyvázání „sociálních vztahů z místních kontextů interakce a jejich restrukturuaci v neomezených časoprostorových rozpětích.“ [Giddens, 2003: 26] Jedním z mechanismů jsou peníze, kterým se v této práci věnovat nebudu a druhým, pro nás důležitým, jsou *expertní systémy*<sup>40</sup>. Pojem expertní systémy používá Giddens pro realizované technické systémy nebo expertizy v rámci kterých je organizováno vědění [ibid: 31] Nejde tak o konkrétní experty, ale o celý systém. Pro většinu laiků není podle Giddense kontakt s odborníky každodenní záležitostí, ale se systémy, ve kterých je expertní vědění koncentrováno jsou všichni v interakci nepřetržitě [ibid.: 32] Jako příklad uvádí život v domě, kdy už jenom přítomností v něm jsme zahrnuti v expertním systému, kdy se nebojíme, že se pod námi zhroutí schody, aniž bychom cokoli věděli o principech stavitelství nebo architektury. Není to o důvěře v konkrétní stavitele, ale v celý abstraktní systém a jeho vědění, jelikož již není v silách jednotlivce být vzdělán ve všech oborech tak, aby vše dokázal kvalifikovaně hodnotit a kontrolovat. Náš svět je totiž takovými systémy protkaný a najdeme je téměř ve všech oblastech lidského života. Nezbývá než jim věřit. [ibid.] Právě důvěra je proto podle Giddense v případě expertních systémů důležitá, jelikož je podmínkou pro fungování celého systému.

Jak je popsáno v kapitole 4.2, důvěra je silně spojena s riziky a jejich vnímáním, kdy funguje jako jeden z mechanismů vyrovnání se s představou rizika. Podle Niklase Luhmanna je důvěra společenským konstruktem k redukci komplexity. [Luhmann, 1979: 29] Také zmiňuje potřebu rozlišení důvěry a důvěřivosti. Důvěra je podle něj spojeno s očekáváním a předpokládá uvědomění si rizika. V jeho pojetí je také vázáno na určitý čas a místo. Důvěřivost je pak pasivní víra v situacích, kdy nám nic jiného nezbývá nebo je pravděpodobnost jejich negativních následků malá. [Luhmann, 1979: 97] Giddens s ním nesouhlasí v tvrzení, že pokud se zdržíme jednání, nevystavíme se riziku. Na něm mu vadí hlavně přístup k riziku, kdy jsou i taková, která vyžadují naše jednání. Také nesouhlasí s takovým popisem důvěry. Ta je podle něj kontinuálním procesem a není vázaná na čas ani prostor. Přidává své vnímání důvěřivosti, kdy důvěřivost vyjadřuje víru v poctivost druhého

---

<sup>40</sup> Expertní systémy je pojem, se kterým přišel Anthony Giddens v rámci popisu jedno ze dvou vyvazujících mechanismů moderní společnosti. Zároveň se ale může plést s oblastí informatiky s identickým názvem, proto je tento typ odlišen dolním indexem jako expertní systém<sub>A1</sub>.

nebo právě přesnost abstraktních principů jako v případě technologií. [Giddens, 2003: 36–39]

Giddens se zamýšlí i přímo nad důvěrou v *abstraktní systémy*. V moderní společnosti již není možné se z těchto systémů vyvázat. Dříve šlo ignorovat prohlášení duchovních vůdců, ale s všudypřítomností expertních systémů to v současnosti již nelze. Důvěra v ně je proto zásadní. Ta za určitých okolností nepředpokládá setkávání s jedinci nebo organizacemi, které jsou za ně odpovědné. [Giddens, 2003: 79] V případě AI to není tak jednoduché. Pokud bereme jako odpovědnou osobu její programátory pak je možné si alespoň existenci někoho konkrétního představit. Situace se ale zkomplikuje, pokud budeme přemýšlet nad AI jako samotným expertem, kterému jen v počátcích zadaly váhy a parametry, ale expertní vědění a informace jako takové dále produkuje sama. Toto dilema nechám otevřené. Giddens mluví o *přístupových bodech* abstraktního systému, to jsou setkání laiků a expertů nebo jejich představitelů. Jsou uznávány z obou zúčastněných stran. Taková setkání zahrnují zapojení laických aktérů a projevují se na nich *obvyklé jednání* ze strany expertů. Ačkoli člověk věří spíše abstraktnímu systému než jeho reprezentantovi, tato setkání nám připomínají lidskost systému, který reprezentují. [Giddens, 2003: 80] Giddens rozděluje na *tvářná* a *beztvářná* setkání. Jak jsem ale nastínila výše, AI jde pod určitým úhlem pohledu brát jako experta a označení tvářné je pak zavádějící. Giddens také v době psaní *Důsledků modernity* v roce 1990 nemohl tušit jakou revoluci v komunikaci způsobí příchod internetové komunikace v reálném čase videokomunikace. V kontextu těchto technologií si myslím, že tvářná nemusí nutně znamenat osobní setkání, ačkoli takové má pro důvěru vyšší hodnotu.<sup>41</sup> Tím pádem do tvářných setkání může pro oblast umělé inteligence spadat jak programátor nebo jiný představitel, tak i samotná AI. Giddens také používá Goffmanovy pojmy *jeviště* a *zákulisi*. Kontrola rozdílu mezi prezentací před a za „oponou“ je důležitá pro obraz profesionality daného oboru. Rozdělení na dvě „scény“ má také význam pro oddělení experta a expertizy. Podle Giddense [Giddens, 2003: 81] „Experti mohou leccos pokazit tím, že to špatně interpretují nebo neznají expertizu, kterou by měli mít k dispozici.“ V lékařském prostředí by tak bylo například těžké slepě důvěřovat lékařům, pokud bychom si byli vědomi všech pochybení, ke kterým v nemocnici dochází. Posledním významem oddělení je, že během fungování abstraktních systémů vždy dochází k nějaké míře neurčitosti. Experti tedy doufají, že laik bude klidnější, pokud mu tuto projev systému neuvidí. [Giddens, 2003: 81]

---

<sup>41</sup> Tématem komunikace zprostředkované počítačem se blíže zabývá Luhmann [Luhmann, 2012: 182–183]

V souvislosti s posledním důvodem pro dělení na jeviště a zákulisí Giddens uvádí následující: „Neexistuje natolik pečlivě osvojená dovednost a natolik obsáhlá forma expertního vědění, která by vyloučila vrtošivost štěstí a náhod.“ Tady bych se opět zastavila nad souvislostí s umělou inteligencí. Pokud jsme si dříve stanovili, že AI je formou expertního vědění a může být i expertem, pak následující tvrzení není tak docela pravdivé. Už z principu fungování jakýkoliv stroj nemůže generovat náhodné výsledky. V případě expertních systémů<sup>AI</sup> člověk zadává vstupní data a požadavky a na tohoto člověka se mohou vztahovat principy náhody a štěstěny. V případě pokročilejších technologií jako například *reinforcement learning*, kdy je AI model postaven před nějakou situací a je mu zadán požadovaný stav<sup>42</sup> Výstup daného modelu umělé inteligence, který jsme neočekávali je tak nahodilý jen v našich očích neznalých kroků, které k tomuto výsledku AI dovedla.

#### 4.1.3 Důvěra ve vědění

*„Proč většina lidí zpravidla důvěřuje praktikám a sociálním mechanismům, o nichž má jen nepatrné, nebo žádné technické vědomosti?“*

[Giddens 2003: 82]

Giddens se touto otázkou zabývá hlouběji. Odpovědí na ni je podle něj nevědomost, jelikož ta se vyskytuje tam, kde je důvěra. Víra v technické vědění je podle něj spojena s pragmatickým postojem skepse k abstraktním systémům. Okolnosti každodenního moderního života ale postoje důvěry vynucují. Jde tak o nevyslovené přijetí okolností bez možnosti alternativní volby. Individuální důvěra nebo její nedostatek bývá silně ovlivněna vědění zprostředkovaným médii nebo jinými zdroji a zkušeností v přístupových bodech, kterým jsem se věnovala v minulé kapitole. V těchto bodech dochází ke střetu laiků a expertů a je tedy náchylné na ohrožení samotného abstraktního systémů v případě negativní zkušenosti. [Giddens, 2003: 84] Podobně říká i Luhmann v souvislosti s důvěrou, která je podle něj zasazena do známého světa pomocí *symbolických reprezentací* a je tedy náchylná na *symbolických událostí*, které mohou její existenci narušit. Jako příklad takové události uvádí zaznamenání situace, která opodstatněnost naší důvěry zpochybní. [Luhmann, 1988: 95–97] Mohlo by se jednat o zveřejněný únik dat jako v případě Cambridge

---

<sup>42</sup> Například postavím AI model před počítačovou hru a zadám ji jako cíl zvyšující se skóre. AI pak bude metodou pokus omyl zkoušet jednotlivé tahy až jejich výsledky budou mít za výsledek zvyšující se skóre.

Analytica<sup>43</sup>. Reakce na takovou zkušenost může mít u méně náročných oborů za následek zvýšení vlastní odbornosti nebo snahu o odpoutání se od systému. [Giddens, 2003: 84] V případě umělé inteligence je pravděpodobnější spíše druhá reakce a ani ta není tak snadná.

Jak vyplývá z předchozích kapitol, umělá inteligence je nyní součástí naše světa, ale i individuálních životů a je tedy těžké se jí vyhnout. V tom je obsažena i nová povaha rizik, jak o ní mluví Beck, kdy téměř není útěku. Zatímco kosmetice vyrobené ze škodlivých látek je možné se vyhnout a pohodlně ji nahradit jinou, u technologií to není tak snadné. Pokud bychom se individuálně rozhodli zavrhnout technologie, znamenalo by to kromě ztráty komfortu i odpojení se od společnosti, bankovního systému, ale i zdravotní péče, infrastruktury... a tak by výčet mohl pokračovat. Luhmann to vidí také pesimističtěji, když říká, že v rámci globální povahy technologických rizik, již není možné před problémy utéct tak, jak velí instinkt. V takovém případě zmiňuje ignorování problému jako nepřekvapivé „řešení“. Jelikož ačkoli je, zřejmé, že nějaká katastrofa v budoucnu nastane, pravděpodobně to nebude zítra. [Luhmann, 1998: 76]

## 4.2 Technika

### 4.2.1 Technika, komunikace a Niklas Luhmann

Tuto část bych ráda věnovala myšlenkám v díle Niklase Luhmanna, jelikož se v jeho díle explicitně objevuje téma počítače, a dokonce i umělé inteligence. Představit celý Luhmannův teoretický aparát a zabývat se jím by bylo na samostatnou práci. Sám Luhmann v předmluvě ke své knize *Sociální systémy: Nárys obecné teorie* popisuje svou teorii jako spíše podobnou „labyrintu, než dálnici vedoucí ke světlým zítřkům.“ [Luhmann, 2006: 11] Proto bych alespo

Raná Luhmannova tvorba byla ovlivněna studijním pobytem na univerzitě v Harvardu u Talcotta Parsonse. Později se však s vydává jinou myšlenkovou cestou a vytváří obecnou teorii moderní společnosti, a to Systémovou teorii. Obecná teorie systémů byla nejdříve formována v první polovině 20. století v oblasti biologie. Teprve poté se stala univerzální vědeckou metodologií užívanou napříč obory, včetně společenských věd. Konkrétně v sociologii pojem systém nahradil do té doby používaný pojem sociální

---

<sup>43</sup> Britská společnost Cambridge Analytica v roce 2014 bez souhlasu použila osobní data k vytvoření systému, který dokázal profilovat jednotlivé americké voliče a nabízet jim personalizovanou politickou reklamu. [Cadwalladar; Graham-Harrison, 2018]

organismus. [Šubrt 2007: 70-71] Jeho dílo bylo významně ovlivněno. Teoretického uvažování o systémech se historicky dělí na tři etapy. Nejprve se objevilo paradigma části a celku, které sahá až k počátku novověku. Takový přístup můžeme nalézt například v díle E. Durheima). Poté se dané paradigma rozšiřuje o odlišení systému a okolí, kterým se zabýval právě Parsons a v počátcích i Luhmann. Ten se však koncem sedmdesátých let od původního pojetí názorově vzdaluje, když pod vlivem nových poznatků kognitivní biologie charakterizuje systémy jako autopoietické<sup>44</sup>. [Šubrt 1996: 27]

Podle Luhmanna je sociální systém tvořen komunikací. Téma komunikace je pro něj důležité; tvrdí, že se zánikem lidského života zanikne i veškerá komunikace a tím nastane i konec společnosti. [Luhmann, 1998: 83] Luhmann převrací klasické pojetí komunikace jako přenosu informací od určitého jedince nebo skupiny k jiným, jako to uvádí například Giddens. [Giddens, 1999: 550] Luhmann nejdříve přebírá tradiční postoj jeho předchůdců, jakým byl například Weber a Parsons, že konstitutivním prvkem sociálních systémů je jednání. Tedy že předpokladem pro sociální systém je, že v jeho rámci spolu více osob smysluplně jedná a je ohraničeno od prostředí k nim nenáležícím. S přijetím nového paradigmatu Luhmann zachovává ohraničenost systému vzhledem k jeho prostředí, ale konstitučním prvkem již nejsou lidé a jejich jednání, ale komunikace. Ke zrodu komunikace v luhmannovském pojetí dochází syntézou tří selekcí a to: 1. informace (kterou někdo vybere z množství jiných), 2. sdělení (vybere způsob pro dané sdělení) a konečně za 3. pochopení [Šubrt, 2007: 87] Pochopení je přeloženo z německého „Verstehen“, které by se dalo přeložit i jako porozumění. To ale v češtině může působit zavádějícím dojmem, jelikož ve třetí fázi komunikace nejde o porozumění významu sdělení nebo konsensu, ale o prostý fakt, že příjemce/ego<sup>45</sup> zaznamenal sdělení a odlišil ho od informace.

Když si představíme náročnost tohoto procesu a téměř nekonečný výsledek různých variací proměnných, vychází z něj komunikace jako značně nejistá až nepravděpodobná. Toho si všímá i Luhmann a nepravděpodobnost označuje jako charakteristický rys komunikačního procesu. Nepravděpodobnost je překonávána v rámci sociokulturní evoluce, která zvyšuje šance na úspěšnou komunikaci. Důležitý je právě pojem evoluce, jelikož se nejedná o prostý proces růstu, ale o proces evoluční a s ním spojený proces selekce, který určí, jaký druh sociálního systému bude možný a který zanikne. Evoluční výhody, které zvyšují pravděpodobnost komunikace nazývá Luhmann médii a jejich účelem je stabilizace

---

<sup>44</sup> Luhman tak označoval systémy, které vznikají a udržují se díky sami sobě.

<sup>45</sup> Luhmann používá pro “odesílatele” “příjemce” pojmy alter a ego [Luhmann, 1995: 80]

komunikačních procesů. Média pak dělí na dva typy podle jejich funkce. Prvním typem jsou média, které umožňují hromadné rozšíření komunikace a umožňují i překonání přirozených překážek komunikace, jako je například prostor. Základním médiem je pro Luhmanna jazyk a k němu se později přidalo i písmo, knihtisk nebo právě elektronická komunikace. Druhým typem jsou média, díky kterým je možné překonat rozdílné motivace během komunikace. Jsou jimi například peníze, láska, moc, víra atd. [ibid.]

Luhmann mimo jiné píše o důležitosti konceptu důvěry v kontextu veškerých sociálních interakcí ve vztahu k moderním technologiím [Luhmann 1979: 48] V Theory of society [2012: 180], například píše, že technická síť je vůči komunikaci neutrální, tedy že informace je produkována mimo danou technickou síť. V telekomunikaci (od té telefonické až po emaily) vidí pouze zprostředkování komunikace dvou stran, ale dle jeho slov nemůže být o nějaké orální nebo psané formě komunikace řeč. [Luhmann, 2012: 181] Na druhou stranu však připouští, že vynález a rozvoj elektroniky pro zpracování informací situací do jisté míry mění. V této souvislosti zmiňuje právě umělou inteligenci a debaty, jestli je schopná čehokoli srovnatelného nebo dokonce nadřazeného lidskému vědomí. Když se zamýšlí nad schopností počítačů nahradit nebo překonat komunikaci, která je pro něj konstitutivním prvkem sociálních systémů, jak bylo řečeno výše, tvrdí, že aby to bylo možné: „museli by zacházet s věděním jako s formou a tedy vědět, co ostatní počítače ne.“<sup>46</sup> [Luhmann, 2012: 181] Odkazuje se zde na letní výzkumný seminář zabývající se umělou inteligencí, který je dnes nazýván jako Dartmouthská konference<sup>47</sup>, kde již bylo dříve řečeno, že umělá inteligence je možná, pokud budeme schopni definovat aspekty lidské inteligence a převést je do podoby kódu. To ale podle něj značí, že umělá inteligence je stále jen o programování, což ji posouvá do oblasti lingvistické komunikace. [Luhmann, 2012: 181] Je nutné říci, že výrok, že umělá inteligence nepřekročí oblast programování není nový. Již v roce 1843, tedy 154 let před Luhmannem se podobně vyjádřila matematicka Ada Lovelace. Ta řekla, že počítače nemohou nic vytvořit, jelikož pouze dělají, co jim člověk skrze programování přikáže. [Bringsjord; Bello; Ferrucci, 2001: 2] Jenže je to tak i nyní, čtvrt století po prvním vydání Die Gesellschaft der Gesellschaft?

V současné době nejpokročilejší systémy využívají neuronové sítě, které jsou ve svém důsledku stále jen řadou matematických funkcí a algoritmů, které tvoří programátoři. Programátor také musí AI dát její účel, například poskytnout dataset nebo iniciovat určitou

---

<sup>46</sup> V originále: They would have to treat knowledge as form, and thus know what other computers do not.

<sup>47</sup> Viz teoretická část.

činnost. Laurence Moroney ze společnosti Google na konferenci Google I/O v roce 2019 popsal rozdíl mezi klasickým programováním a strojovým učením: „Při tradičním programování zadáváte pravidla a data a program vám poskytne odpovědi. Ve strojovém učení zadáváte data a odpovědi a program vypisuje pravidla.” [Moroney, 2019]

Program tedy zatím nemůže vytvořit nic mimo rámec toho, pro co byl stvořen. Vždy je limitován tím, co do něj programátoři vložili. Existují pokročilé datové modely pro vytváření “nových” obrázků ze zadaného textového popisu<sup>48</sup> či pro vytváření “nových” textových útvarů<sup>49</sup>, avšak stále se jedná jen o manipulaci existujících zdrojových dat a chování, pro která byl program vytvořen. Často používaný argument, že neuronová síť je “černá skříňka” o které ani vývojáři nevědí, jak funguje, reflektuje pouze to, že vývojáři nevidí, proč program přisuzuje určitou důležitost jednotlivým transformacím, nikoliv to, že neví, co se v něm děje. Umělá inteligence je tedy stále o programování, jelikož i u nejpokročilejších technologií je alespoň v počátku nutný lidský zásah.

Dále Luhmann zmiňuje nenahraditelnost a nadřazenost orální a psané komunikace. Tu obhajuje tím, že komunikace je probíhající zpracovávání rozdílů mezi věděním a nevěděním, aniž by bylo nutné zjišťovat, jaké (ne)vědomosti jednotlivci nebo stroje mají. Dodává ale, že tyto typy komunikace mohou využít počítače k posílení vlastních schopností a koncentrovat se na podstatné věci, které nemohou být nahrazeny technologií. [Luhmann, 2012: 181] V tomto bodě je nutné dá Luhmannovi za pravdu, jelikož orální a psaná komunikace je stále nadřazená té počítačem vytvořené, ale díky technice je nyní možné ji zprostředkovat na mnohem delší vzdálenosti (jako v případě různých platforem jako je například Skype, MS Teams a jiných) nebo překonávat i jazykovou bariéru v reálném čase.<sup>50</sup> Existují například různá AI řešení pro osoby s poruchami řeči, zraku nebo sluchu, kdy jim umělá inteligence umožňuje lepší komunikaci a tím i vyšší kvalitu života a začlenění do společnosti. Zatímco překlady „text-to-text“ v jiných jazycích v reálném čase jsou již běžně dostupné<sup>51</sup>, nyní jsou ve fázi výzkumu a testování i „speech-to-speech“ překlady. Takovým je projekt opět od technologického giganta Google “Translatron” z roku 2019 a jeho vylepšená verze „Translatron 2” z roku 2021. Druhá verze tohoto programu, díky kterému

---

<sup>48</sup> Například program DALL-E viz <https://openai.com/dall-e-2/>.

<sup>49</sup> Například GPT-2 viz <https://openai.com/blog/better-language-models/>.

<sup>50</sup> O transformaci času a prostoru mluví i Anthony Giddens v *Důsledcích modernity*. [Giddens, 2003: 23-26], ale také Ulrich Beck [Beck, 2018:].

<sup>51</sup> Například aplikace Google Lens, díky které stačí zamířit telefonem na text v cizím jazyce a na obrazovce se v reálném čase objevuje text již přeložený.



překlad z jedné řeči do druhé zněl vaším hlasem (na rozdíl od toho elektronického se kterým je možné se setkat nyní) obsahoval nejen technologická vylepšení, ale především se snažil vypořádat s obavou ze zneužití v rámci deepfakes. [Jia; Ramanovich; Remez; Pomerantz, 2021] Právě s touto oblastí měl Luhmann při představách o budoucnosti s počítači problém. Na příkladu videa jako média pro komunikaci zmiňoval jeho závislost na reálném čase. V této souvislosti přímo řekl: „Tato závislost na reálném čase produkuje určitý bonus důvěryhodnosti: není tu žádný čas na komplexní manipulaci nebo zjišťování takové manipulace ať už během natáčení nebo sledování.“<sup>52</sup> [Luhmann, 2012: 183]

#### 4.2.2 Technika bez člověka

Abychom se dostali k problému nerovností, musíme se vrátit zpátky do počátečních předpokladů vědy. Dualitou věd se zabývá například Bruno Latour v knize *The Politics of Nature*. Tam bádá po vzniku tradičního rozdělení světa na přírodu a kulturu a tuto tradici nachází až v antickém Řecku. Pro popis používá Platónovo podobenství o jeskyni, kde skupina – společnost sedí přikovaná uvnitř a ze stínů, které na zeď vrhá oheň jsou pro ně jediným známým světem. Pro osvětlení pravdy je potřeba postava filosofa – vědce, který dokáže zpravovat osoby o pravdě tam venku. Latour, který se zabýval studiem vědy, používá k vysvětlení vztahu přírody a kultury. Kdy kultura je uvnitř jeskyně a jen se dohaduje, kdežto venku je to skutečné a exaktní – příroda. Jeden svět venku a jeden uvnitř. [Latour, 2004]

Jedna z nejdůležitějších otázek naší společnosti je: Jak vlastně chceme žít? A na ni není možné odpovědět bez zahrnutí společenských věd do debat o na první pohled technických otázkách, jelikož při konstatování rizik je podle Becka dosud nerozvinutou symbiózou přírodních a společenských věd, racionality každodenního žití a racionality expertů. Nemohou být od sebe odděleny a rozvíjeny na základě vlastních racionalit [Beck, 2018: 37] Pokud k tomuto přesto dochází, vznikají rozdílné a často protikladné definice a boje o ně. V prostředí takovýchto bojů přichází věda o svůj monopol na racionalitu, jelikož vždy budou existovat rozdílné a navzájem si konkurující zájmy, a bude odkázána na sociální očekávání.

O specifickém postavení přírodních věd mluví také Beck, když říká, že v rámci problémů životního prostředí, které se týkají nás všech se diskuze vedou stále v kategoriích a formulích přírodních věd. To je samozřejmě platné i pro formulaci rizik plynoucích

---

<sup>52</sup> V originále: „But this dependence on real time produces a certain credibility bonus: there is no time for complex manipulation or to check for such manipulation either during shooting or viewing.“

z umělé inteligence. Říká [Beck, 2018: 31]: „Zůstává při tom nerozpoznáno, že přírodovědné „formule zbídačení“ mají sociální, kulturní a politický význam.“ A nejdou tedy vést bez zahrnutí společenských věd. Beck se dokonce vyjadřuje i ke konkrétnímu problému techniky a průmyslu [ibid.]: „Právě diskuse posledních desetiletí, během nichž byl opět jednou předesťřen a prezentován celý arzenál argumentů kritizujících techniku a průmysl, zůstaly v zásadě technokratické a naturalistické.“ A dodává [ibid.]: „Tomuto myšlení [...] zůstávají od počátku skryty sociální, politické a kulturní obsahy a důsledky rizik spojených s modernizací.“ A také odpověď na otázku, jak vlastně chceme žít. A na ní není možné odpovědět bez zahrnutí společenských věd do debat o na první pohled technických otázkách, jelikož „konstatace rizik jsou dosud nerozpoznanou, nerozvinutou symbiózou přírodních a duchovních věd, racionality každodenního života a racionality expertů, zájmu a faktu. Nejsou současně ani jen jedním, ani jen druhým. Jsou obojím, a to v nové formě. Nemohou už být v duchu specializace izolovány od jednoho či druhého a rozvíjeny a specifikovány na základě standardů vlastní racionality“ [Beck, 2018: 37] Pokud k tomuto přesto dochází, vznikají rozdílné a často protikladné definice a boje o ně. V prostředí takovýchto bojů přichází věda o svůj monopol na racionalitu, jelikož vždy budou existovat rozdílné a navzájem si konkurující zájmy, a bude odkázána na sociální očekávání.

### 4.2.3 Člověk není proměnná

Beck upozorňuje na *ztrátu sociálního myšlení*, pokud uvažujeme bez sociálního kontextu. Situace, ve kterých se snažíme dopad, ohrožení nebo riziko nějakého jevu zredukovat na „průměrné“, jenomže společnost není homogenní. Stejně tak v situaci dvou lidí, kdy jeden má dvě jablka nemůžeme říci, že oba mají průměrně jedno jablko, když jeden z nich hladoví. Zajímavé podle něj také je, že toto aplikování matematických principů není nápadná často ani sociologům. [Beck, 2018: 32-33] Beck dále myšlenku rozvádí: „uvnitř vystavenosti rizikům existují nové *sociální nerovnosti*. Ty vznikají především tam, kde se – rovněž v mezinárodním měřítku – překrývají třídní a rizikové situace: proletariát světové rizikové společnosti sídlí pod továrními komíny, vedle rafinerií a chemických továren v průmyslových centrech třetího světa.“ Do těchto oblastí se s oblibou zbavují rizik ti, kteří mají moc o tom rozhodnout. Zmiňuje také systémovou přitažlivost mezi krajní chudobou a krajním rizikem, pro kterou mluví i to, že si nezaměstnané obyvatelstvo raději vybírá neviditelné ohrožení plynoucí z nových technologií před o oči bijící hmotnou nouzí. [Beck, 2018: 55] Zapomínáme tak ve prospěch jedné strany na miliony lidí bez přístupu k základním lidským potřebám, natož pak moderním technologiím. Ti se s ní setkají až na

posledním stupni jejího životního cyklu, kdy už pozbyla všech benefitů a zbyly jen zdroje ohrožení.<sup>53</sup> Stejně jako příkladu s jablky, kdy nikdo nemá jablek stejně, ale jeden hladoví a druhý má naopak více než potřebuje.

Téma nerovnosti je ale hluboce zakořeněné i do umělé inteligence samotné. Jak již bylo nastíněno v teoretické části, jedním z vážných a aktuálních problémů umělé inteligence je *code bias* tedy předpojatost kódu, kdy se díky nekvalitě a zastaralosti dat tvořených na podkladu sociálních nerovností tato předpojatost reprodukuje dál.

Code bias je problém nejen etický ale i společenský. Klasické přístupy k AI, například expertní systémy AI tímto problémem mohli trpět také, avšak je mnohem znatelnější u strojového učení. Rozhodnutí učiněná umělou inteligencí mohou tak být nespravedlivá a mohou prohlubovat již existující problém předpojatosti vůči určitým etnikám nebo skupinám osob. Code bias je v drtivé většině případů neúmyslný. Velmi často vzniká z nedostatečné znalosti, jak systémy AI fungují, z nevědomosti o této technologii nebo z vlastní nevědomé předpojatosti. [Coeckelbergh, 2020: 71-72] Ve světě, kde je spousta rozhodnutí vykonáno s doporučením softwaru nebo pouze přímo jím, code bias přináší závažné důsledky. Jednotlivcům může být upřen přístup ke zdrojům a svobodám, mohou nedostat úvěr nebo skončit ve vězení jen kvůli svému etniku. Rozhodnutí nemusí ovlivnit pouze jednotlivce, ale celé skupiny nebo obyvatele určité části města, kterou AI vyhodnotilo jako rizikovou. [ibid.]

Konkrétní případy můžeme nalézt po celém světě. Na začátku roku 2020 rozhodl soud v Holandském Haagu o zákazu používání softwaru, který vyhodnocoval osoby žijící v chudších částech země jako ty s větší pravděpodobností spáchání podvodu se sociálními dávkami. Pro tento účel program SyRI shromažďoval 17 kategorií dat včetně daňových záznamů a registračních značek automobilů. Soud rozhodl, že shromažďování takových dat bez souhlasu porušuje GDPR a co více, že předurčením každého, kdo žije v nesprávné oblasti, jako osoby s vyšší pravděpodobností spáchání trestného činu je porušení lidských práv. [Burack, 2020]

Již v roce 2011 poukázali Nick Bostrom a Eliezer Yudkowsky na možný problém “z blízké budoucnosti”, kdy žadatel o hypotéku může být odmítnut na základě rozhodnutí AI a žalovat banku, že toto rozhodnutí bylo diskriminační kvůli rase žadatele. Banka by se bránila

---

<sup>53</sup> Podle studie deníku Environmental Science and Technology bylo v roce 23 % elektronického odpadu ze zemích globálního Severu vyvezeno do sedmi zemí globálního Jihu. [Breivik; Armitage; Wania; Jones, 2014]

argumentem, že to není možné a že algoritmus je “slepý” k rase žadatele a nestranný (což by pravděpodobně pro banku byl i hlavní důvod na implementaci takového algoritmu). [Bostrom, Yudkowsky, 2011] Jak by tedy bylo možné, že algoritmus odmítal afroamerické uchazeče? Pokud by byla základem použitého algoritmu složitá neuronová síť nebo genetický algoritmus, bylo by velmi složité či téměř nemožné zjistit proč se tak děje a jak je vůbec schopný rozlišit rasu uchazeče. V případě použití založeném na rozhodovacích stromech nebo bayesovských sítích je proces učení transparentnější a programátor by dokázal zjistit, že algoritmus používá informace o adrese žadatelů, kteří se narodili nebo dříve bydleli v převážně chudých oblastech. Tato blízká budoucnost se stala realitou ani ne o 10 let později, kdy ke skutečnému odmítání uchazečů na základě bydliště docházelo. Server The Markup v rozsáhlé studii zjistil, že věřitelé v roce 2019 častěji odmítali půjčky na bydlení lidem jiné než bílé etnicity s podobnými finančními charakteristikami. [Martinez, Kirchner, 2021] Při komplexní statistické analýze více než dvou milionů žádostí o konvenční hypotéku na koupi domu server zjistil, že věřitelé se 40% vyšší pravděpodobností odmítají půjčky žadatelům z řad Latinoameričanů, s 50% vyšší pravděpodobností odmítají půjčky žadatelům z řad Asiátů/Tichomořských ostrovů a s 70% vyšší pravděpodobností odmítají půjčky žadatelům z řad původních obyvatel USA a dokonce 80 % věřitelů častěji odmítlo afroamerické žadatele oproti žadatelům z řad majoritního etnika. Tyto hodnoty byly naměřené v rámci celé země. Ve všech případech vypadali neúspěšní žadatelé o půjčku “na papíře” téměř stejně jako žadatelé bílé etnicity, až na výjimku jejich rasy. [ibid.]

Oficiálně o půjčkách rozhoduje úředník, reálně však algoritmy. V Americe nejčastěji ty od polovládních organizací Freddie Mac a Fannie Mae – algoritmy vyvinuté na základě údajů z 90. let a více než 15 let staré. Je všeobecně považován předpojatý/biased vůči menšinám, protože odměňuje tradiční úvěry, k nimž má bílá populace lepší přístup. Nezohledňuje mimo jiné včasné platby za nájem, komunální služby a účty za mobilní telefon, avšak snižuje skóre lidí, kteří se s nimi dostanou do prodlení a je na ně uvalena exekuce. Na rozdíl od novějších modelů penalizuje lidi za dluhy za zdravotní péči z minulosti, i když byly splaceny. [ibid.] Jeden z výzkumů [Rice, Swesnik, 2013] také ukázal, že poskytovatelé krátkodobých půjček obvykle umísťují své pobočky do čtvrtí, kde žijí převážně obyvatelé etnických minorit a kde je méně poboček bank. V důsledku toho je pravděpodobnější, že obyvatelé využijí těchto služeb k půjčení peněz mnohem vyšší. To vytváří neúplnou a zkreslenou úvěrovou historii, protože banky úvěrovým úřadům hlásí jak dobré, tak špatné finanční chování, zatímco poskytovatelé krátkodobých půjček hlásí pouze zmeškané splátky.

Nejen barva pleti může být znevýhodněna AI. Obdobný problém vedl dokonce k vyřazení interního náborového systému ve společnosti Amazon v roce 2018, který diskriminoval ženy [Dastin, 2018]. Systém se učil na datech, která velmi často obsahovala technické pozice, na které se častěji hlásili muži a pohlaví tak bylo algoritmem zvoleno jako jeden z parametrů pro výběr CV na přijetí nebo na postup do dalších kol výběrového řízení. Ženám dával záporné body pouze na základě pohlaví. Pohlaví bylo rozlišováno na základě školy nebo pomocí použití ženského tvaru v textu.

V tak rizikovém prostředí, jako je zdravotní péče je předpojatost obzvláště znepokojivá. Navíc nejvíce dopadá na již diskriminované skupiny zdravotním systémem – jak ze strukturálních faktorů (horší dostupnost), tak z vědeckých (studie, výzkum a postup léčby vznikly ve spojitosti s jinou skupinou obyvatelstva). Již v současnosti se systémy umělé inteligence postavené na strojovém učení používají ke stanovení individuálních pooperačních plánů k léčbě bolesti či k předpovídání rakoviny. Také hrají roli při rozhodování o distribuci orgánů, vývoji vakcín a dalších prvků zdravotní péče. Předpojatost v těchto rozhodnutích může mít bez nadsázky charakter života a smrti.

V roce 2019 vědci zjistili, že v USA široce používaný algoritmus Optum, který předpovídá, kteří pacienti budou mít prospěch z nadstandardní lékařské péče, dramaticky podhodnocuje zdravotní potřeby nejnemocnějších afroamerických pacientů, což jen prohlubuje dlouhodobé rasové rozdíly v medicíně. Algoritmus byl naprogramován tak, aby výslovně vyloučil rasu, a přesto dokázal najít způsob, jak identifikovat převážně afroamerické pacienty. Použil zdánlivě rasově neutrální metriku, kolik by pacienti v budoucnu stáli zdravotnický systém. Toto měřítko však ve skutečnosti rasově neutrálním měřítkem není jelikož afroameričtí pacienti vynaložili ročně na zdravotní péči přibližně o 1800 dolarů méně než bílí pacienti se stejným počtem chronických onemocnění. Algoritmus tedy vyhodnotil jako rovnocenné „středně“ nemocné pacienty bílé pleti s těžce nemocnými Afroameričany. [Johnson, 2019] Pokud by se tato předpojatost v algoritmu neobjevila, zdvojnásobil by se počet afroamerických pacientů označených jako ohrožených komplikovaným zdravotním stavem. Když společnost Optum zopakovala analýzu na celostátním souboru dat 3,7 milionu pacientů, zjistila, že afroameričtí pacienti, které algoritmus vyhodnotil stejně rizikové jako bílé pacienty, byli mnohem nemocnější – dohromady u nich bylo zjištěno o 48 772 chronických onemocnění víc. [ibid.]

Nedávná studie také ukázala, že nový algoritmus k identifikaci melanomů je přesnější než lékař. Avšak riziko, že tento algoritmus bude předpojatý k afroamerickým pacientům se jeví jako velmi vysoké. Algoritmus byl trénován převážně na skupině obyvatel

světlé pleti. [Lashbrook, 2018] Správná diagnóza této skupiny obyvatel je důležitá i z hlediska dožití. Pětiletá míra přežití je o Afroameričanů o 17 % nižší, než u bílého obyvatelstva USA. Tato technologie má velký potenciál zachránit spoustu životů a generovat rychlé diagnózy, avšak opět se zde naráží na problém kvality dat. Nesprávně naučený algoritmus může způsobit větší škody, jelikož dá pacientům falešný pocit bezpečí nebo nesprávně diagnostikuje rakovinu.

Nové algoritmy se dostávají na trh rychleji než dříve a hlásají lékařské zázraky. Ztráta povědomí o zakořeněné předpojatosti ve zdrojových datech by mohla vést k odepření adekvátní péče minoritním skupinám obyvatel, které nebyly zahrnuty do jejich vytváření. Ačkoliv neznáme dlouhodobé dopady využití umělé inteligence ve zdravotnictví, je zde riziko, že celkový pokrok pro většinovou populaci může jít na úkor těch nejzranitelnějších.

Beck jako řešení nabízí následující [Beck, 2011: 293]: “Jestliže nemají být vedlejší účinky déle tolerovány, musí vědecko-technický vývoj při svém tempu a formách, jimiž prochází, zaručit v každém stadiu schopnost učení se. To předpokládá, že se bude vyhýbat procesům vytvářejícím ireverzibilní situace.” Také přidává radu se zaměřovat na souvislosti a tím tedy i více spolupráce mezi společenskou a přírodní vědou. Jelikož kdo jiný by mohl s otázkami společnosti pomoci lépe, než společenská věda.

## 5 Diskuze

V této části práci bych se ráda věnovala diskuzi výsledků analýzy a zodpovězení na výzkumné otázky. Dokázala společenská věda předjímat vývoj budoucí inteligence a je tedy nosná i v případě této oblasti?

Nejprve bych se chtěla věnovat vybranému typu metody pro zodpovězení si otázek mého cíle. Pokud jde o zjištění odpovědi na samotnou otázku týkající se možnosti predikce budoucího vývoje společenské vědy pro umělou inteligenci, pak hodnotím zvolené metody jako správně vybrané. Jelikož jak jsem přestřela na začátku práce, kdy jsem se rozhodovala mezi několika typy textové analýzy, zajímala mne skutečná výpovědní hodnota tvrzení vybraných autorů v případě umělé inteligence v testu času. Proto by mne například frekvenční analýza dle mého názoru omezovala ve volnosti při procházení textů na předem stanovené kategorie pojmů. To se v případě Althusserovy metody symptomálního čtení nedělo. Metodu jsem zvolila přesto, že je pro závěrečnou práci spíše nezvyklá. Myslím, že právě díky ní se mi z textem lépe pracovalo a byla jsem schopna nacházet souvislosti, které bych jinak neviděla. Byla bych nejspíš svázána vlastním diskurzem tak, jak to píše ve své knize *Čist Kapital* sám Althusser. Další vybraná metoda hermeneutiky, kterou jsem používala jako pomocnou, mi dávala pevnější rámec pro mou práci, což se osvědčilo hlavně v počátcích psaní. Jsem spíše analytický typ a díky hermeneutice jsem měla v počátcích stanovený alespoň rámcový postup pro práci s literaturou. Zároveň ale „spirálový“ a nejen kruhový postup hermeneutiky působil sám o sobě inspirativně, kdy jsem společenskovědní texty četla po stále vyšších úrovních a byla schopná vnímat obsahy sdělení v konkrétním textu.

### 5.1 Výzkumné otázky

Nyní se budu věnovat zodpovězení hlavních otázek a podotázek výzkumu, které byly následující:

- Dokáží společenské vědy anticipovat budoucnost v případě umělé inteligence?
  - Kde se ne/objevují jednotlivé pojmy v různých typech literatury?
  - Jak se liší pojetí rizika v jednotlivých typech literatury?
  - V jakých kontextech se jednotlivé pojmy objevují?
  - S jakými důsledky se jednotlivé pojmy vyskytují?
  - Jsou témata, která jeden nebo druhý typ literatury nepopisuje?

- Jak se liší jazyky a obsahy obou typů popisu (v hlavních rysech)?

Odpověď na hlavní otázku si nechám na konec. Nyní se zaměřím na podotázky. Nejdříve je třeba podotknout, že jsem si je stanovila na začátku práce možná také díky svému technickému myšlení, kdy jsem chtěla mít zadání a výsledky své práce více „pod kontrolou“. Pak mne ale díky zvoleným metodám výzkumu nesla témata prací více organicky. Nyní, v závěru si myslím, že takto položené otázky by byly vhodnější při použití zmíněné frekvenční analýzy textu. Nicméně se pokusím odpovědět na otázky tak, jak jsou položeny.

*Kde se ne/objevují jednotlivé pojmy v různých typech literatury?*

Zrovna tato otázka spadá pod ty, které by byly lépe zodpovězeny v rámci frekvenční analýzy. Byla však formována v době, kdy jsem předpokládala, že si na začátku definuji seznam pojmů, které chci analyzovat. Nakonec jsem analýze dala volnější řád a nechala se jejími výsledky vést.

*Jak se liší pojetí rizika v jednotlivých typech literatury?*

Texty z oblasti umělé inteligence odpovídaly tvrzení společenskovední literatury, která upozorňovala a varovala před technickým pojetím rizik. V tom se k riziku přistupuje jako k výslednici pravděpodobnosti dané negativní události a rozsahu škod, proti čemu se vyhrazuje Luhmann [1998] nebo se zvažuje poměr benefitů vůči negativním následkům a bere se jako předpoklad, že pokud si správně nastavíme všechny proměnné, bude možné riziku předejít, jak například říká Beck v rámci „kalkulovatelnosti rizik“ [Beck,2011: 289]. V rámci analyzovaného díla společenskovedních autorů se o riziku mluví hodně v jeho důsledcích a charakteristikách a vyhýbají se definicím ve formě: „riziko je..“ následované podstatnými jmény, v případě společenských vědců jsou to spíše přídavná jména. Beck o něm mluví v souvislosti s jeho proměnnou v moderní, rizikové společnosti. Popisuje jeho askriptivní úděl, všudypřítomnost a jeho nárůstu oproti tradičním společnostem. [Beck,2011] Giddens [2000] se nejvíce z vybraných autorů věnuje definici a historii rizika a čerpají z něj i ostatní autoři. Luhmann [1993] podobně jako Giddens vysvětluje vývoj vnímání rizika a shoduje se s oběma autory, že riziko je neodmyslitelně spjata s budoucností, jelikož tam se odehrávají potenciální negativní události [Beck, 2011], [Giddens, 2000] [Luhmann, 1993].

*V jakých kontextech se jednotlivé pojmy objevují?*

Jelikož se otázka vztahuje na pojmy, je na ni těžké přesně odpovědět z důvodu popsání v první otázce. Nicméně bych chtěla rozvést kontext rizika, jakožto jednoho z pojmů, který jsem původně chtěla zkoumat. Na tuto otázku jsem částečně odpověděla v té



předchozí, jen bych dodala, že texty přímo řešící AI je dávaly do podobného kontextu, jaký zmiňovala společenskovední literatura v souvislosti s předpoklady vědy. U většiny z nich tedy problém rizik vyzníval tak, že pro mitigaci rizika je potřeba si důsledně stanovit všechny parametry a bádát více. To byl i poznatek Becka [2011] a Giddense [2000] v důsledku víry ve vědu a moc poznání. Tedy že je potřeba ještě více vědy a výzkumu. Zároveň se i v oblasti AI vyskytuje spousta skeptiků hlásajících zdrženlivost v technologiích umělé inteligence nebo jejich úplné upuštění. Zástupcem, obvykle z řad futuristů, by mohl být Bostrom nebo Tegmark, podle nichž jsme na rozcestí, jestli AI bude tou nejlepší věcí, která se lidstvu stala nebo tou poslední. Podobné názory sdíleli i jiní autoři, ale nedostali takový prostor v rámci mého výběru, jelikož po definování aktuálních hrozeb AI, které vyplývají z teoretické části práce jsem se nezabývala takzvanou obecnou inteligencí nebo superinteligencí.

*S jakými důsledky se jednotlivé pojmy vyskytují?*

Odpověď je stejná jako na první otázku.

*Jsou témata, která jeden nebo druhý typ literatury nepopisuje?*

Obecně bych řekla, že povětšinou ne. Oblast umělé inteligence je všudypřítomná a řeší tedy rozličná témata a všichni vybraní autoři společenských věd byli vybráni pro svůj holistický pohled na svět. Poslední jmenovaní jsou pochopitelně limitováni časovým kontextem, ale vždy se alespoň jeden z autorů zabýval tématem, které vystalo z druhého typu literatury. Co bych ale ráda zmínila je vývoj mého postoje k této otázce v čase. Znovu musím potvrdit vhodnost výběru metody pro výzkum. Jelikož čím hlouběji jsem postupovala v rámci symptomálního čtení a hermeneutiky, tím více se přede mnou otevíral „význam za textem“.

*Jak se liší jazyky a obsahy obou typů popisu (v hlavních rysech)?*

Tato otázka by se také lépe zodpovídala při užití jiné metody, například jazykové analýzy, frekvenční analýzy a jiných. Nicméně co se týče jazyka, literatura týkající se AI a zvláště pak studie, používali autoři technické popisy a strohý jazyk. To se měnilo u populárně naučné literatury, kdy je dán důraz nejen na jasnost obsahu i pro ne odborníky, ale i na čtivost textu.

Společenskovední autoři používali jazyk obvyklý pro tento typ literatury. Z hlediska náročnosti jazyka vynikal Luhmann, který zvláště v díle *Sociální systémy* byl i vzhledem k pestré škále vlastních pojmů často hůře uchopitelný. Z důvodu téměř neexistence jeho děl v češtině jsem jeho díla četla v angličtině a některé jeho obraty vyžadovaly opakované přečtení. Beck psal *Rizikovou společnost* v jazyce, který měl být dostupný širokému obecnstvu, ale v českém překladu, který jsem měla k dispozici, vyznívaly jeho obraty občas

těžkopádně. Nejlépe se četl Giddens, kdy bez zbytečných lingvistických obrátů sděloval co měl na mysli, ale bez ztrát na obsahu sdělení. Je také možné, že jak Beck, tak i Luhmann psali v němčině, a proto zmíněné připomínky mohly být způsobeny překladem do českého nebo německého jazyka.

U obou typů literatury bylo specifické pojmosloví. Zatímco v AI literatuře je spousta pojmů ryze technických, troufám si říci, že většina z nich se v českém odborném prostředí používá v původním tvaru jako samotná „AI“, „deepfakes“, „bias“ nebo „deep learning“. To má podle mě původ v prostředí, kde tyto práce vznikají a pro koho jsou určeny. To, jak zažité jsou anglikanismy v těchto kruzích pramení i z jazyka, v jakém se programuje, tvoří a debatuje. Je to však i důkazem tvrzení všech tří společenskovedních autorů o globalizaci. Virtuální svět AI a počítačové komunikace je už bytostně globalizovaný a sdílí tedy i jeden společný jazyk. V rámci společenskovední literatury zase každý autor přichází se svým pojmoslovím, částečně i proto, aby své myšlenky a teorie odlišil od ostatních.

Nyní se dostávám k samotné hlavní otázce, a to: „*Dokáží společenské vědy anticipovat budoucnost v případě umělé inteligence?*“

Na tuto otázku bych se nejprve chtěla podívat v kontextu samotného předvídání. Pokud bychom brali předvídání z přírodovědného pohledu, potřebovali bychom přesně dané časové osy, trendy, pravidla a výchozí stavy proměnných. O to mé v práci nejde. Proto účelově hledat jeden konkrétní výrok, vztah nebo stav by nemělo smysl. Ono „předvídání budoucnosti“ má tak spíše charakter zachování výpovědní hodnoty pod testem času a v mém případě aplikování na technologii, která byla v době psaní textů ještě v počátcích. Vzhledem k velikosti oblastí jako je AI a společenské vědy jako obor není možné porovnat všechny různé texty a přístupy. A ani to k zodpovězení otázky není nutné. Pokud použiji příklad filosofa Karla Poppera s černými labutěmi – pokud bychom chtěli prokázat, že všechny labutě nejsou bílé, pak stačí jedna černá k potvrzení hypotézy. Nemusíme je hledat všechny. I proto se v analýze objevuje pouze krátký výčet nalezených témat.

K samotné schopnosti sociologie předvídat budoucnost byl kritický i sám Niklas Luhmann. Kritizuje klasickou kritiku společnosti<sup>54</sup>, místo toho, aby se sociologie více zaměřila na důsledky technologií a jejich rizik a nebezpečí. Sociologie se dle jeho názoru přestala ptát na jednu ze svých základních otázek a to: „Co za tím vším je?“ [Luhmann, 1998,

---

<sup>54</sup> Dokonce i přímo zmiňuje Ulricha Becka

77] Ačkoli se mu to v době psaní jeho textu nemuselo zdát, moje práce přichází s jiným výsledkem.

V první části práce jsem si stanovila několik oblastí AI, které jsem vyhodnotila jako rizikové, a proto jsem se na ně rámcově zaměřila i v části analýzy.

Ve společenskovední literatuře se tomuto tématu věnoval nejvíce Ulrich Beck. Jeho *Riziková společnost* věnuje kapitolu vědě, jejímu vzniku a problematickým předpokladům. Přichází s pojmem *scientizace*, tedy zvědečt'ování, které dělí na dvě etapy: prostou a reflexivní. Během prosté scientizace, spojované s neochvějnou vírou v pokrok a vědu, je laik konfrontován s neotřesitelnou pozicí experta. [Beck, 2004: 257] Právě v tomto vztahu může mít původ i jedna z nejistot v této oblasti, a to laika. Umělá inteligence zvláště na svém počátku vyžadovala vysoce expertní vědění zahrnující hned několik oblastí vědy, a to umožnilo formovat hradbu za kterou zůstala veřejnost. Beck v pasáži o vědě tvoří až děsivě přesné popisy o feudalizaci poznávací praxe. Zmiňuje pozitivní stránku, díky které můžeme být méně závislí na autoritním pohledu na vědu. Luhmann stejnou myšlenku formuluje neuvěřitelně přesně na to, v jaké fázi vývoje byl Internet v roce 1997, kdy kniha vyšla poprvé. Říká [Luhmann, 2012: 187]: „V zásadě každý si bude v budoucnu moci zkontrolovat výroky expertů, jako jsou například lékaři a právníci, na svém počítači.“ Což je, řekla bych, jedna z nejpřesnějších a nejkonkrétnějších predikcí, které jsem v literatuře našla. To ale mohlo být způsobeno i větším zájmem Luhmanna o oblast informatiky, než měli jeho kolegové. Díky tomu mohlo být snazší odušit vývojový trend.

Beck zároveň zmiňuje, jakým způsobem se může dostupnost informací vymknout, pokud nás vnitřní rozpory vědy a její postoj k laikům donutí se od ní odklonit a zhlédneme se ve vědění *nových alchymistů*. [Beck, 2011: 278] To souvisí s druhým problémem, který jsem si identifikovala. Informací je totiž k nalezení nepřeberné množství a argument lze nalézt pro jakákoli tvrzení. Díky téměř neomezeným informacím je však pro mnohé lidi těžké ověřit spolehlivost zdroje. Umělá inteligence na tomto má i negativní podíl, kdy nejenže díky absenci kontextu nedokáže rozlišit to dobré (ale to ani člověk ne), ale zároveň díky ní lze vytvářet zcela alternativní podklady pro živá tvrzení jako v případě deepfakes. Zároveň je také díky algoritmům schopná vytvořit kolem uživatele tzv. sociální bublinu, kdy jsou jejím členům podsouvány články, v lepším případě díky zájmu o konkrétní oblast a v tom horším jako úmyslná manipulace s veřejným míněním. Díky všem těmto aspektům je v této době téměř neomezeného přístupu k informacím paradoxně těžší si utvořit kritický názor.

Ačkoli Giddens v roce 1990 při psaní *Důsledků modernity* mohl jen stěží tušit o rozvoji na poli umělé inteligence, právě ta je, dle mého názoru, skvělým příkladem expertního systému, přestože poněkud nestandardním. Může být formou systému koncentrujícího vědění například ve formě různých expertních systémů<sup>AI</sup> sloužících k diagnostice ve zdravotnictví, tak, jak o něm mluví Giddens výše. AI ve smyslu systému vědění jsme v dnešním světě doslova obklopeni. Spoléháme na ni, když otiskem prstu autorizujeme platbu v rámci mobilního bankovníctví, pijeme vodu z úpravny vod, kde její kvalitu řídí systémy s umělou inteligencí, spoléháme na ní v rámci chytrého řízení dopravy. Umělá inteligence je také specifická tím, jak v relativně krátké době byla schopna doslova prorůst jinými obory v rámci expertního systému vědění a tím fungování celého systému ještě více utužuje. Pokud si tak dříve člověk dokázal alespoň představit, že dům, na jehož dobrou statiku spoléhá tvořili lidé-odborníci, pak s příchodem umělé inteligence do nejrůznějších oborů se takovému člověku kauzalita ještě více vzdálila. I daní odborníci si totiž chodí pro rady k dalšímu odborníkovi, a to samotné AI. Tím je dle mého názoru umělá inteligence v kontextu expertních systémů unikátní.

Poslední bod, který bych chtěla zmínit code bias a s ní spojenou reprodukce nerovností. Problematikou nerovností se zabývá jak Beck tak i Giddens<sup>55</sup>, kteří řeší jak téma nerovností v rámci světa, tak i v rámci společností. Luhmann je naopak kritizován kvůli konzervativnímu přístupu a opomíjení nerovností ve společnosti<sup>56</sup>. Ani jeden z autorů se přímo nezabýval tím, jak samotná technika může přispět k dalšímu prohlubování nerovnosti v rámci etnik, genderu a různých minoritních skupin. Zároveň ale Beckovy výroky o důležitosti zahrnutí společenských věd i do oblastí věd přírodovědných, v kontextu problematiky AI, vyznívají nadčasově.

Pokud bychom se zaměřili i na špatné odhady, pak bych zmínila ten Giddensův, že [Giddens, 2003: 81]: „Neexistuje natolik pečlivě osvojená dovednost a natolik obsáhlá forma expertního vědění, která by vyloučila vrtošivost štěstí a náhod.“ Pokud chápeme AI jako formu expertního vědění, pak toto tvrzení nemůže být pravdivé, jelikož už z podstaty fungování stroje není možné generovat náhodné výsledky.

Další výrok pronesl Niklas Luhmann, když na příkladu videa jako média pro komunikaci zmiňoval jeho závislost na reálném čase. V této souvislosti přímo řekl: „Tato závislost na reálném čase produkuje určitý bonus důvěryhodnosti: „není tu žádný čas na

---

<sup>55</sup> Dále téma rozpracovává v knize *Třetí cesta* z roku 1998, která nebyla předmětem analýzy

<sup>56</sup> Hlavně z pohledu nerovnosti pohlaví.

komplexní manipulaci nebo zjišťování takové manipulace at' už během natáčení nebo sledování.<sup>57</sup> [Luhmann, 2012: 183] To ale odporuje jednomu z největších současných rizik AI týkající se manipulaci obrazové formy tzv. *deepfakes*, která může ohrožovat i demokracii.

Pokud mám tedy konkrétněji odpovědět na původní otázku: „*Dokáží společenské vědy anticipovat budoucnost v případě umělé inteligence?*“ Pak má odpověď zní **ano**. Všichni tři autoři se ve svých dílech zabývali alespoň jednou z oblastí rizik AI, které jsem na začátku práce identifikovala. Pokud jde o jejich individuální míru relevance pro toto téma pak bych začala Niklasem Luhmannem, který se tématu počítačů, a dokonce i umělé inteligenci věnoval explicitně. Jeho představy (at' už se naplnily nebo ne) byly nejpřesnější. Anthony Giddens se zabýval podobným pohledem na společnost jako Ulrich Beck a skvěle zpracoval téma expertního vědění. Jeho poznatky o globalizaci a vymknutí společnosti z času a prostoru by byly ještě nosnější, kdybych se problematice AI věnovala například z čistě technického nebo ekonomického hlediska. Posledním z autorů byl Ulrich Beck, na kterého lze najít názory, že za jeho popularitu může z části skvělé načasování<sup>58</sup>. V knize jsou i problematice pasáže, kdy si v rámci několika stran protirečí, když jednou říká, že rizika jsou demokratizační a o kousek dál, že rizika prohlubují nerovnosti. Je také výrazně orientovaný na prostředí Německa a Evropy. Nicméně ze všech autorů nabídl nejucelenější pohled na problematiku rizik moderní společnosti a s jeho závěrem, že je nutné propojit svět přírodních a společenských věd nelze než souhlasit.

## 5.2 Výběr autorů

Zhodnotit výběr autorů k analýze, pak si myslím, že pro danou práci byli vybráni správně. Každý z nich ve svém díle reflektuje témata ekologie a dává ji do vztahu se společností. I to byl předpoklad holistického pohledu na vybrané téma, který se mi potvrdil.

Pokud bych nyní měla volit znovu, pravděpodobně bych se zaměřila pouze na jednoho autora a na něj se zaměřila detailněji, řešila i celkový kontext jeho tvorby, doby a místa kde žil. Díky tomu by bylo možné jeho myšlenky pochopit hlouběji.

---

<sup>57</sup> V originále: „But this dependence on real time produces a certain credibility bonus: there is no time for complex manipulation or to check for such manipulation either during shooting or viewing.“

<sup>58</sup> Kniha *Riziková společnost*, ve které se objevuje silný motiv varování před nukleární havárií byla ve fázi příprav, když vybuchla jaderná elektrárna v Černobylu.

Pokud bych měla napsat ještě třetí práci, vybrala bych si autory tvořící před druhou světovou válkou, což bylo jedno z kritérií volby literatury k analýze. Díky studiu tématu jsem se dostala k dílu Martina Heideggera, Josého Ortegy y Gasset a Arnolda Gehlena. Už jenom při zběžném prolistování byly jejich myšlenky týkající se techniky neskutečně přesné a jsem přesvědčena, že pokud bych se těmito autory zabývala detailněji, pak jejich teze o technice budou v případě umělé inteligence mnohem přesnější a detailnější. To je dle mého názoru způsobeno tím, že v době, kdy tvořili se mohli nerušeně zamýšlet nad všemi možnými formami, dovednostmi a důsledky techniky, aniž by byli ovlivněni konkrétními vynálezy, které vlivem války a pokroku v průmyslu po nich. To se projevuje i na dílech autorů, které jsem nakonec zvolila. Ti řeší hlavně řekněme chemicko-fyzikální povahu rizik a u Becka je z díla cítit naléhavost rizik nukleární energie.

## 6 Závěr

Tato diplomová práce se od počátku soustředila na výpovědní hodnotu společenských věd v případě technologií. Cílem bylo zjistit, co společenské vědy dokázaly predikovat v případě zdánlivě technické a exaktní oblasti, jako je umělá inteligence.

Práce měla také za cíl zmapovat vývoj umělé inteligence a identifikovat její reálné současné hrozby. Co se těch týče, vybrala jsem tři aktuální, které mi přišly zásadní. První takovou byla celková nejasnost, co se týče definic, co vlastně AI je a není a co by mohla být. Na počátku umělé inteligence jako oboru byla v roce 1956 konference na Dartmouth College, který si dal za cíl tuto oblast dále rozvíjet. V roce 2006 byla uspořádána další konference k výročí padesáti let a její organizátor James Moor řekl: „Přestože umělá inteligence zaznamenala za posledních 50 let mnoho úspěchů, přetrvává v této oblasti řada dramatických sporů. Různé oblasti výzkumu často nespolupracují, výzkumníci používají různé metodologie a stále neexistuje obecná teorie inteligence nebo učení, která by obor sjednocovala.“<sup>59</sup> To naznačuje, že v této oblasti panuje nejistota dodnes. Co se týče povědomí lidí o AI, stále je velice ovlivněno chytlavými titulky a kulturními zobrazeními. Právě to může vést k tomu, že nynější rozvoj AI opět skončí další zimou anebo se začne používat pro tuto oblast informatiky jiný název, který bude mít přesně definovanější obrysy. Díky pokroku v oblasti a open source komunitě se ale k veřejnosti v poslední době dostávají praktické aplikace této technologie, se kterou můžou interagovat. a tím se může zkonkrétnit představa o umělé inteligence a tím v důsledku i věcnějším debatám veřejnosti a expertů ohledně AI. O tom, jak se vztah veřejnosti a expertů psal Giddens, který se zabýval důvěrou laiků v abstraktní systémy. Na počátku kapitoly si klade otázku, jak lidé mohou věřit něčemu, o čem mají pramalé technické povědomí. Odpovědí mu bylo, že jim díky nevědomosti nic jiného nezbyvá. Právě tady by mohlo pomoci tuto oblast více demystifikovat a přiblížit lidem.

Druhým problémem byla schopnost tvořit texty a obrázky tak přesvědčivé, že je již dnes těžké je rozeznat od reality<sup>60</sup>. Díky tomu se otvírá možnost manipulovat názory ve společnosti a vrážet do ní klíny v době, kdy jak jsme se všichni během posledních dvou let přesvědčili, jsme na sobě závislejší čím dál víc. Ulrich Beck poskytl vhled do fungování

---

<sup>59</sup> MOOR, James. 2006. The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years. 27. vyd.: American Association for Artificial Intelligence. ISBN: 0738-4602-2006

<sup>60</sup> To se ostatně můžete přesvědčit v Příloze.

vědy a jejích vnitřních rozkolů, které v důsledku vytvářejí prostor pro různé *nové alchymisty*, jak je sám nazval. Právě ti mohou technologii AI využívat ke špatným účelům.

Poslední oblastí byla další reprodukce společenských nerovností díky zastaralým kódům s vnitřní předpojatostí. Ty jsou způsobeny nejen neaktuálností dat, ale i předpojatostmi ve společnosti, ze které vzešly. Podle Luhmanna je komunikace základním konstitutivním prvkem v sociálním systému a bál se, jaký bude počítačem zprostředkovaná komunikace mít vliv na společnost. Nové technologie tak můžeme využít na prohlubování nerovnosti a nebo je využít jako skvělého pomocníka k překonání bariéry času a prostoru a komunikovat spolu navzájem napříč kontinenty, vrstvami a skupinami a najít společnou řeč.

V této práci jsem se zabývala současnými hrozbami. Jak ale bylo patrné na začátku mého úvodu a snad vyznělo i z celé mé práce, oblast umělé inteligence je technologií rozvíjející se mílovými skoky. Je tedy nasnadě jestli (spíše kdy) bude vyvinuta obecná umělá inteligence a superinteligence a jak dlouho s ní budeme zvládat držet krok. Této oblasti jsem se úmyslně nevěnovala pro její jiné obecné předpoklady a další výzkum by bylo možné vést tímto směrem.

Nevědomost, důvěra a předpojatost. To jsou pojmy, které by krátce shrnuly vážné problémy současné AI ohrožující člověka. Ne nedostatek čipů nebo pomalý kód. A s těmito problémy dokáže pomoci společenská věda. Cílem mého výzkumu bylo zjistit, zda společenské vědy dokáží anticipovat budoucnost a prokázala jsem, že ano. Co je na tom ale nejdůležitější je opětovné potvrzení významu společenské vědy na vývoji, výzkumu v rámci věd přírodovědných. Jelikož stejně jako není člověka bez přírody a přírody bez člověka, neměla by být ani přírodní věda bez té společenskovední.



## 7 Seznam použité literatury

### 7.1 Metodologická literatura

ALTHUSSER Luis; BALIBAR Étienne. 1970. *Reading capital*. London: New Left Books. Dostupné z: <https://www.marxists.org/reference/archive/althusser/1968/reading-capital/index.htm> 21. 6. 2022 15:37:39

GRONDIN, Jean. 1997. *Úvod do hermeneutiky*. Z něm. orig. přel. Horyna, Břetislav; Kouba, Pavel. 1. vyd. Praha: Oikoymenh. 247 s. ISBN 80-86005-43-7

HAUSER Michael. 2011. *Althusserovo symptomální čtení jako lekce historického materialismu*. Filosofía: nakladatelství Filosofického ústavu AV ČR. 406 s. ISBN:9788070074176

HENDL, Jan. 2005. *Kvalitativní výzkum. Základní teorie, metody a aplikace*. 2. vyd. Praha: Portál. 408 s. ISBN 978-80-7367-485-4

KUŽEL, Petr. 2014. *Althusser: filosofie a politika. o filosofii, která chce změnit svět*. Dostupné z: <https://sok.bz/clanky/2012/petr-kuzel-althusser-filosofie-a-politika-o-filosofii-ktera-chce-zmenit-svet> 21. 6. 2022 15:40:13

### 7.2 Literatura

ANDERSON, Duncan. 2021. *When AI writes poetry*. Humanise. Dostupné z: <https://humanise.ai/blog/ai-writes-poetry/> 7.6. 2022 12:50:59

BAJWA, J., MUNIR, U., NORI, A., WILLIAMS, B. (2021). Artificial intelligence in healthcare: transforming the practice of medicine. *Future healthcare journal*, 8(2), e188–e194. Dostupné z: <https://doi.org/10.7861/fhj.2021-0095> 17. 6. 2022 21:38:42

BARNETT, Edward. 2019. *Generative Adversarial Networks: Revitalizing old video game textures*. Towards Data Science. Dostupné z: <https://towardsdatascience.com/generative-adversarial-networks-revitalizing-old-video-game-textures-669493f883a0> 18. 6. 2022 20:37:57

BATEMAN, Tom. 2021. *France approves fully autonomous bus for driving on public roads in a European first*. euronew. Dostupné z: <https://www.euronews.com/next/2021/12/01/france-approves-fully-autonomous-bus-for-driving-on-public-roads-in-a-european-first> 11. 6. 2022 14:02:28

BECK, ULRICH. 2018. *Riziková společnost (Na cestě k jiné moderně)*. 3. vyd. Praha: Sociologické nakladatelství (SLON). 431 s. ISBN: 978–80–7419–267–8

BECK, Ulrich. 2018. *Riziková společnost: Na cestě k jiné modernitě*. 3. vyd. Praha: Sociologické nakladatelství. 431 s. ISBN: 978-80-7419-267-8

BECK, ULRICH; BECK–GERNSHEIM, Elizabeth. 2002. *Individualization: Institutionalized Individualism and its Social and political Consequences*. London: Sage Publications. 221 s. ISBN 0–7619–6111–9

BECK, Ulrich. 2007. *Vynalézání politiky*. 1. vyd. Praha: Slon. 273 s. ISBN: 978-80-86429-64-9

BODEN, Margaret A. 2021. *AI: its nature and future*. 1. vyd. Oxford: Oxford University Press. 198 s. ISBN: 9780191083488 0191083488

BOSTROM, Nick. 2017. *Superintelligence: Superintelligence: Až budou stroje chytřejší než lidé*. 1. vyd. Praha: Prostor. 512 s. ISBN 978-80-7260-353-4

BOSTROM, Nick; YUDKOWSKY, Eliezer. 2021. *The ethics of artificial intelligence*. Dostupné z: <https://nickbostrom.com/ethics/artificial-intelligence.pdf> 18. 6. 2022 19:18:41

BREIVIK Knut, ARMITAGE James M., WANIA Frank, JONES Kevin C. 2014. *Tracking the Global Generation and Exports of e-Waste. Do Existing Estimates Add up?*. ACS Publications. Dostupné z: <https://pubs.acs.org/doi/10.1021/es5021313> 16. 5. 2022 18:47:33

BREIVIK, Knut; James M. ARMITAGE; Frank WANIA; JONES, Kevin C. 2014. *Tracking the Global Generation and Exports of e-Waste. Do Existing Estimates Add up?*. American Chemical Society (ACS). Dostupné z: <https://pubs.acs.org/doi/10.1021/es5021313> 2. 5. 2022 19:50:26

BRINGSJORD, Selmer; BELLO, Paul; FERRUCCI, David. 2001. *Creativity, the Turing Test, and the (Better) Lovelace Test*. *Minds and Machines* 11, 3–27, Dostupné z: <https://doi.org/10.1023/A:1011206622741>

BROWN, Jennings. 2019. *Go Champion Retires After Realizing AI Is 'an Entity That Cannot Be Defeated'*. Gizmodo. Dostupné z: <https://gizmodo.com/go-champion-realizes-ai-has-rendered-his-life-meaningless-1840068855> 19. 6. 2022 21:25:46

BROWNLEE, Jason. 2019. *A Gentle Introduction to Generative Adversarial Networks (GANs)*. Machine Learning Mastery. Dostupné z: <https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/> 21. 6. 2022 0:23:37

BURACK, Justin. 2020. *Addressing Algorithmic Discrimination in the European Union*. A Path for Europe. Dostupné z: <https://pathforeurope.eu/addressing-algorithmic-discrimination-in-the-european-union/> 26. 6. 2022 22:01:42

CADWALLARD, Carole; GRAHAM-HARRISON, Emma. 2018. *Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach*. The Guardian. Dostupné z: <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election> 7. 6. 2022 14:38:40

CARPENTER, Bob. 2021. *EU proposing to regulate the use of Bayesian estimation*. Statistical Modeling, Causal Inference, and Social Science. Dostupné z: <https://statmodeling.stat.columbia.edu/2021/04/22/eu-proposing-to-regulate-the-use-of-bayesian-estimation/> 21. 6. 2022 10:43:56

COECKELBERGH, Mark. 2020. *AI ethics*. 1. vyd. Cambridge: MIT Press. 248 s. ISBN: 9780262357067 0262357062

CUTHBERTSON, Anthony. 2022. *'The Game is Over': Google's DeepMind says it is on verge of achieving human-level AI*. Independent. Dostupné z: <https://www.independent.co.uk/tech/ai-deepmind-artificial-general-intelligence-b2080740.html> 23. 6. 2022 21:22:39

DASTIN, Jeffrey. 2018. *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. Dostupné z: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G> 6. 6. 2022 18:35:45

DEEPMIND. 2016. *Deep Reinforcement Learning*. Deepmind. Dostupné z: <https://www.deepmind.com/blog/deep-reinforcement-learning> 11. 6. 2022 21:41:42

DEEPMIND. 2020. *AlphaFold: a solution to a 50-year-old grand challenge in biology*. Deepmind. Dostupné z: <https://www.deepmind.com/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology> 24. 6. 2022 14:22:01

DEEPMIND. 2022. *A Generalist Agent*. Deepmind. Dostupné z: <https://www.deepmind.com/publications/a-generalist-agent> 7. 6. 2022 9:34:28

EWING, Jack. 2021. *How Germany Hopes to Get the Edge in Driverless Technology*. New York Times. Dostupné z: <https://www.nytimes.com/2021/07/14/business/germany-autonomous-driving-new-law.html> 7. 6. 2022 18:16:53

FOOTE, Keith D. 2022. *A Brief History of Deep Learning*. Dataversity. Dostupné z: <https://www.dataiversity.net/brief-history-deep-learning/> 23. 5. 2022 12:34:13

FORD, Martin. 2011. *Will IBM's Watson put your job in jeopardy?*. Fortune. Dostupné z: <https://fortune.com/2011/02/15/will-ibms-watson-put-your-job-in-jeopardy/> 24. 5. 2022 12:00:40

FRANKISH, Keith; RAMSEY, William M. 2018. *The Cambridge handbook of artificial intelligence*. 1. vyd. Cambridge: Cambridge University Press. 354 s. ISBN: 9781139046855

GAMMON, Katharine. 2011. *Watson Goes to the Hospital*. Technology Review. Dostupné z: <https://www.technologyreview.com/2011/02/23/24086/watson-goes-to-the-hospital/> 12. 6. 2022 0:57:48

GARTNER. 2021. *Gartner Identifies Four Trends Driving Near-Term Artificial Intelligence Innovation*. Gartner. Dostupné z: <https://www.gartner.com/en/newsroom/press-releases/2021-09-07-gartner-identifies-four-trends-driving-near-term-artificial-intelligence-innovation> 19. 6. 2022 0:46:08

GEROVITCH, Slava. (2011), *The Search for a Theory of Cognition: Artificial Intelligence With A National Face: American And Soviet Cultural Metaphors For Thought*

GESSNER, D., Mar 1, 2021, 5:59 PM, "The real reasons we don't have flying cars", online, dostupné z: <https://www.businessinsider.com/why-flying-cars-still-dont-exist-2021-2>

GIDDENS, Anthony. 1999. *Sociologie*. 1. vyd. Praha: Argo. 596 s. ISBN: 80-7203-124-4

GIDDENS, Anthony. 2000. *Unikající svět: jak globalizace mění náš život*. 1. vyd. Praha: Slon. 135 s. ISBN 80-85850-91-5.

GREENE, Tristan. 2022. *DeepMind's astounding new 'Gato' AI makes me fear humans will never achieve AGI*. The Next Web. Dostupné z: <https://thenextweb.com/news/deepminds-astounding-new-gato-ai-makes-fear-humans-will-never-achieve-agi> 18. 6. 2022 16:12:50

GROENFELDT, Tom. 2022. *IBM Puts AI On A Chip To Improve Fraud Detection In Real-Time Payments*. Forbes. Dostupné z: <https://www.forbes.com/sites/tomgroenfeldt/2022/04/06/ibm-puts-ai-on-a-chip-to-improve-fraud-detection-in-real-time-payments/> 23. 6. 2022 23:04:03

GROH, Matthew, EPSTEIN, Ziv., FIRESTONE, Chaz., PICARD, Rosalind. 2021. *Deepfake detection by human crowds, machines, and machine-informed crowds*. Forbes. Dostupné z: <https://www.pnas.org/doi/full/10.1073/pnas.2110013119> 5. 6. 2022 16:57:39

GUSTIN, Sam. 2011. *IBM Watson Scientist: Speed Matters, But So Do Accuracy, Intuition*. Wired. Dostupné z: <https://www.wired.com/2011/02/ibm-watson-speed/> 30. 5. 2022 17:55:59

HEAVEN, Will D. 2022. *This horse-riding astronaut is a milestone in AI's journey to make sense of the world*. Technology Review. Dostupné z: <https://www.technologyreview.com/2022/04/06/1049061/dalle-openai-gpt3-ai-agi-multimodal-image-generation/> 25. 6. 2022 13:00:26

CHRISTIE'S. 2018. *Is artificial intelligence set to become art's next medium?*. Christie's. Dostupné z: [https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx?sc\\_lang=en](https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx?sc_lang=en) 18. 6. 2022 18:57:00

IBM. 2012. *A Computer Called Watson*. IBM. Dostupné z: <https://www.ibm.com/ibm/history/ibm100/us/en/icons/watson/> 16. 6. 2022 22:52:27

IBM. 2021. *IBM Watson asistent pomáhá Air Bank zvyšovat kvalitu péče o zákazníky*. IBM Newsroom. Dostupné z: <https://cz.newsroom.ibm.com/2021-06-16-IBM-Watson-asistent-pomaha-Air-Bank-zvysovat-kvalitu-pece-o-zakazniky> 3. 6. 2022 15:19:32

JENNINGS, Ken. 2019. *The Jeopardy! Minor Leagues*. Slate. Dostupné z: <https://slate.com/culture/2019/04/jeopardy-quiz-bowl-connection-ken-jennings.html> 8. 6. 2022 22:53:14

JIA, YE; RAMANOVICH, Michelle T.; REMEZ, Tal, POMERANTZ, Roi. *Translatotron 2: High-quality direct speech-to-speech translation with voice preservation*. Dostupné z: <https://arxiv.org/pdf/2107.08661.pdf>

JOHNSON, Carolyn Y. 2019. *Racial bias in a medical algorithm favors white patients over sicker black patients*. The Washington Post. Dostupné z: <https://www.washingtonpost.com/health/2019/10/24/racial-bias-medical-algorithm-favors-white-patients-over-sicker-black-patients/> 17. 6. 2022 0:15:05

KAPLAN, Andreas. 2020. *Artificial Intelligence, Socialmedia, And Fake News: Is This The End Of Democracy?*. Dostupné z: <https://cdn.istanbul.edu.tr/file/JTA6CLJ8T5/DF3BF2DBD76E4F7DB8A7143D0EA6F212> 4. 6. 2022 13:13:28

KASÍK, Pavel. 2022. „Někdo, ne něco.“ *Vývojář věří, že se naplnil sci-fi scénář*. Seznamzpravy.cz. dostupné z: <https://www.seznamzpravy.cz/clanek/tech-technologie-prvni-robot-jako-clovek-umela-inteligence-ma-dusi-tvrdi-programator-goo-glu-205928> 26. 6. 2022 20:59:49

KELLY, Kevin. 2014. *The Three Breakthroughs That Have Finally Unleashed AI on the World*. Wired. Dostupné z: <https://www.wired.com/2014/10/future-of-artificial-intelligence/> 3. 6. 2022 10:21:58

KOCIJAN, Vid., DAVIS, Ernest., LUKASIEWICZA, Thomas., MARCUS, Gary., MORGENSTERNE Leora. 2022. *The Defeat of the Winograd Schema Challenge*. . Dostupné z: <https://doi.org/10.48550/arXiv.2201.02387> 22. 6. 2022 13:06:44

KUČERA, Roman. 2017. *The truth behind Facebook AI inventing a new language*. Towards Data Science. Dostupné z: <https://towardsdatascience.com/the-truth-behind-facebook-ai-inventing-a-new-language-37c5d680e5a7> 3. 6. 2022 13:00:45

KUHN, Thomas S. 1997. *Struktura vědeckých revolucí*. 1. vyd. Praha: Oikoymenh. 206 s. ISBN: 9788086005546

KUKLOVÁ, Daniela. 2022. *Platforma pro AI modely pomáhá identifikovat rizikové faktory pro detekci rakoviny prsu*. Trask. Dostupné z: <https://www.thetrask.com/post/platforma-pro-ai-modely-pomaha-identifikovat-rizikove-faktory-pro-detekci-rakoviny-prsu> 22. 6. 2022 17:30:00

KURZWEIL, Ray. 2018. *The singularity is near: when humans transcend biology*. 1. vyd. Richmond Duckworth. 652 s. ISBN: 9780715635612

KUŽEL, Petr. 2014. *Filosofie Louise Althussera: o filosofii, která chtěla změnit svět*. 1. vyd. Praha: Filosofía, ISBN 9788070074176

LASHBROOK, Angela. 2018. *AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind*. The Atlantic. Dostupné z: <https://www.thetrask.com/post/platforma-pro-ai-modely-pomaha-identifikovat-rizikove-faktory-pro-detekci-rakoviny-prsu> 20. 6. 2022 12:09:18

LATOUR, Bruno. 2009. *Politics of Nature: How to Bring the Sciences into Democracy*. 1. vyd. Cambridge: Harvard University Press. 320 s. ISBN: 9780674012899

LEONDES, Cornelius T. 2002. *Expert systems : the technology of knowledge management and decision making for the 21st century*. 6. vyd. San Diego: Academic Press. 1947 s. ISBN: 9780124438804

LEVESQUE, Hector J. 2014. *On our best behaviour*. ScienceDirect. Dostupné z: <https://sciencedirect.com/science/article/pii/S0004370214000356> 21. 6. 2022 18:41:30

LUHMANN, Niklas. 1979. *Trust and Powers*. Chichester: John Wiley & Sons. 208 s. ISBN: 9780471997580

- LUHMANN, Niklas. 1988. *Familiarity, Confidence, In Trust: Problems and Alternatives*. vyd. Oxford: Blackwell. 147 s. ISBN: 9780804732345
- LUHMANN, Niklas. 1993. *Risk: A Sociological Theory*. . vyd. Berlin - New York: Walter de Gruyter. 236 s. ISBN: SBN 3-11-012941 -8
- LUHMANN, Niklas. 1998. *Observations on Modernity*. vyd. Stanford, CA: Stanford University Press. 147 s. ISBN: 9780804732345
- MARCUS, Gary; DAVIS, E. 2020. *GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about*. Technology Review. Dostupné z: <https://www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion/> 5. 6. 2022 20:06:13
- MARKOFF, John. 2011. *Computer Wins on 'Jeopardy!': Trivial, It's Not*. New York Times. Dostupné z: <https://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html> 26. 6. 2022 18:56:14
- MARQUORDT, Christian. 2020. *Autonomous minibus in Monheim*. Urban Transport Magazine. Dostupné z: <https://www.urban-transport-magazine.com/en/autonomous-minibus-in-monheim/> 14. 6. 2022 13:59:04
- MARTINEZ, Emmanuel; KIRCHNER, Lauren. 2021. *The Secret Bias Hidden in Mortgage-Approval Algorithms*. The Markup. Dostupné z: <https://themarkup.org/denied/2021/08/25/the-secret-bias-hidden-in-mortgage-approval-algorithms> 20. 6. 2022 18:08:55
- MAŘÍKOVÁ, Hana; PETRUSEK, Miloslav; VODÁKOVÁ, Alena. 1996. *Velký sociologický slovník*. vyd 1. Praha: Karolinum. 1459 s. ISBN: 80-7184-311-3
- MEYER, Bertrand. 2011. *John McCarthy*. Communications of the ACM. Dostupné z: <https://themarkup.org/denied/2021/08/25/the-secret-bias-hidden-in-mortgage-approval-algorithms> 7. 6. 2022 15:21:17
- MITCHELL, Thomas M. 1997. *Velký sociologický slovník*. 1. vyd. Singapore : McGraw-Hill. 352 s. ISBN: 9780070428072
- MORONEY, Laurence. 2019. *Learning Zero to Hero (Google I/O'19)*. YouTube. Dostupné z: <https://www.youtube.com/watch?v=VwVg9jCtqaU> 7. 6. 2022 14:55:19
- MORRISON, Alan. 2022. *Ten years of Google Knowledge Graph*. Data Science Central. Dostupné z: <https://www.datasciencecentral.com/ten-years-of-google-knowledge-graph/> 19. 6. 2022 21:06:13
- NARANG, Sharan; CHOWDHERY, Sharan. 2022. *Pathways Language Model (PaLM): Scaling to 540 Billion Parameters for Breakthrough Performance*. Google AI

Blog. Dostupné z: <https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html> 25. 5. 2022 14:45:08

NÁVRH NAŘÍZENÍ EVROPSKÉHO PARLAMENTU A RADY, KTERÝM SE STANOVÍ HARMONIZOVANÁ PRAVIDLA PRO UMĚLOU INTELIGENCI (AKT O UMĚLÉ INTELIGENCI) A MĚNÍ URČITÉ LEGISLATIVNÍ AKTY UNIE, V Bruselu dne 21.4.2021, dostupné z: <https://eur-lex.europa.eu/legal-content/CS/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>

NEWBORN, M., LIESERSON, Charles E. 2003. *Deep Blue: an artificial intelligence milestone*. 1. vyd. New York: Springer. 346 s. ISBN: 9780387954615.

NEWQUIST, Harvey P. 1994. *The Brain Makers: Genius, Ego, And Greed in the Quest For Machines That Think*. 1. vyd. Indianapolis, Ind: Sams Publishing. 488 s. ISBN: 9780672304125

NIEKERK van, Mariette; SUBRAMANIAN, Aravind. 2022. *How AI is Shaping the Future of Financial Services*. Deloitte. Dostupné z: <https://www2.deloitte.com/nz/en/blog/financial-services/2022/ai-in-financial-services.html> 23. 6. 2022 15:57:53

NILSSON, Nils J. 2013. *The Quest for Artificial Intelligence*. 1. vyd. Cambridge: Cambridge University Press. 562 s. ISBN: 9780521116398

OHNSMAN, Alan. 2022. *Waymo Starts Fully Autonomous Rides In San Francisco, Expanding Arizona Robotaxi Zone*. Forbes. Dostupné z: <https://www.forbes.com/sites/alanoohnsman/2022/03/30/waymo-starts-fully-autonomous-tests-in-san-francisco-expanding-arizona-robotaxi-zone/?sh=6c1d28b95502> 22. 5. 2022 11:12:43

PARK, Terri. 2021. *Behind Covid-19 vaccine development*. MIT News. Dostupné z: <https://news.mit.edu/2021/behind-covid-19-vaccine-development-0518> 21. 6. 2022 14:59:11

PEARSON, Jordan. 2014. *Forget Turing, the Lovelace Test Has a Better Shot at Spotting AI*. Vice. Dostupné z: <https://www.vice.com/en/article/pgaany/forget-turing-the-lovelace-test-has-a-better-shot-at-spotting-ai> 13. 5. 2022 21:48:16

PIPER, Kelsey. 2020. *GPT-3, explained: This new language AI is uncanny, funny — and a big deal*. Vox. Dostupné z: <https://www.vox.com/future-perfect/21355768/gpt-3-ai-openai-turing-test-language> 15. 6. 2022 20:23:57

POSTMAN, Neil. 1993. *Technopoly: the Surrender of Culture to Technology*. 1. vyd. New York: Random House US. 145 s. ISBN: 9780679745402



- PROCHÁZKA, David. 2011. *GPT-3, explained: This new language AI is uncanny, funny — and a big deal*. DSL.cz. Dostupné z: <https://www.dsl.cz/clanky/2419-retro-vite-jak-vypadal-internet-v-roce-1997> 12. 5. 2022 10:16:12
- REDMON, Joseph . 2021. *I stopped doing CV research....* Twitter. Dostupné z: <https://twitter.com/pjreddie/status/1230524770350817280> 11. 5. 2022 21:23:35
- RICE, Lisa; SWESNIK, Deidre. 2014. *Discriminatory Effects of Credit Scoring on Communities of Color*. Dostupné z: [https://cpb-us-e1.wpmucdn.com/sites.suffolk.edu/dist/3/1172/files/2014/01/Rice-Swesnik\\_Lead.pdf](https://cpb-us-e1.wpmucdn.com/sites.suffolk.edu/dist/3/1172/files/2014/01/Rice-Swesnik_Lead.pdf) 31. 5. 2022 16:24:57
- RIEDL, Mark O. 2014. *The Lovelace 2.0 Test of Artificial Creativity and Intelligence*. . Dostupné z: <https://arxiv.org/pdf/1410.6142.pdf> 2. 6. 2022 10:20:13
- ROMPORTL, Jan. 2022. *Podívali jsme se pod kapotu umělé inteligence*. PPF. Dostupné z: <https://www.ppf.eu/newsroom/podivali-jsme-se-pod-kapotu-umele-inteligence-s-janem-romportlem> 6. 5. 2022 12:10:14
- RUSSEL, Stuart. NORVIG, Peter. 2020. *Artificial Intelligence: A Modern Approach, 4th Edition*. 1. vyd. London: Pearson Education. 2145 s. ISBN: 9780134610993
- SHEN, T., 2018 Deep Fakes ” using Generative Adversarial Networks ( GAN ), online, dostupné z: [http://noiselab.ucsd.edu/ECE228\\_2018/Reports/Report16.pdf](http://noiselab.ucsd.edu/ECE228_2018/Reports/Report16.pdf)
- SHEN, Tianxiang a kol. 2018. „Deep Fakes” using Generative Adversarial Networks (GAN), dostupné z: [http://noiselab.ucsd.edu/ECE228\\_2018/Reports/Report16.pdf](http://noiselab.ucsd.edu/ECE228_2018/Reports/Report16.pdf)
- SHILEDARBAXI, Nikita. 2011. *Hands-on Guide To GANSynth: An Adversarial Neural Audio Synthesis Technique*. Analytics India Magazine. Dostupné z: <https://analyticsindiamag.com/hands-on-guide-to-gansynth-an-adversarial-neural-audio-synthesis-technique/> 4. 6. 2022 18:20:08
- SCHUCHMANN, Sebastian. 2019. *History of the Second AI Winter*. Towards Data Science. Dostupné z: <https://towardsdatascience.com/history-of-the-second-ai-winter-406f18789d45> 7. 6. 2022 13:15:58
- SIMON, Julien. 2017. *Fascinating Tales of a Strange Tomorrow*. Towards Data Science. Dostupné z: <https://towardsdatascience.com/fascinating-tales-of-a-strange-tomorrow-72048639e754> 27. 5. 2022 18:56:18
- ŠUBRT, Jiří. 1996. *Luhmannův teoretický pokus o postižení komplexity moderní společnosti*. In Acta Universitatis Carolinae. Philosophica et historica. 2.vyd. ISSN 0567-8293. Dostupné z: doi:10.14712/24647055.2018.60 21. 6. 2022 18:41:30

TEGMARK, Max. 2017. *Life 3.0: Being Human In The Age Of Artificial Intelligence*. 1. vyd. New York: Alfred A. Knopf. 364 s. ISBN: 9781101946596

TEMPLETON, Brad. 2021. *Baidu Launches Robotaxi Service In Outer Beijing. How 'Fully Driverless' Is It Vs. AutoX?*. Forbes. Dostupné z: <https://www.forbes.com/sites/bradtempleton/2021/04/29/baidu-launches-robotaxi-service-outside-beijing--how-fully-driverless-is-it-vs-autox/?sh=46f8b88d6a05> 29. 6. 2022 18:38:18

TORRES, Jordi. 2020. *A gentle introduction to Deep Reinforcement Learning*. Towards Data Science. Dostupné z: <https://towardsdatascience.com/drl-01-a-gentle-introduction-to-deep-reinforcement-learning-405b79866b44> 4. 6. 2022 0:48:42

TORRES-HUITZIL, Cesar. 2017. *Fault and Error Tolerance in Neural Networks: A Review*. Dostupné z: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8013784> 21. 5. 2022 11:03:15

TZABAN, Rotem, MOKADY, Ron., GAL, Rinon, BERMANO, Amit, COHEN-OR Daniel, 2022. *Stitch it in Time: GAN-Based Facial Editing of Real Videos*. Dostupné z: <https://stitch-time.github.io/> 25. 6. 2022 0:57:08

VICTOR, Daniel. 2015. *Elon Musk and Stephen Hawking Among Hundreds to Urge Ban on Military Robots*. New York Times. Dostupné z: <https://www.nytimes.com/2015/07/28/technology/elon-musk-and-stephen-hawking-among-hundreds-to-urge-ban-on-military-robots.html> 13. 5. 2022 17:00:20

VINCENT, J., Mar 5, 2019, "A never-ending stream of AI art goes up for auction", online, dostupné z: <https://www.theverge.com/2019/3/5/18251267/ai-art-gans-mario-klingemann-auction-sothebys-technology>

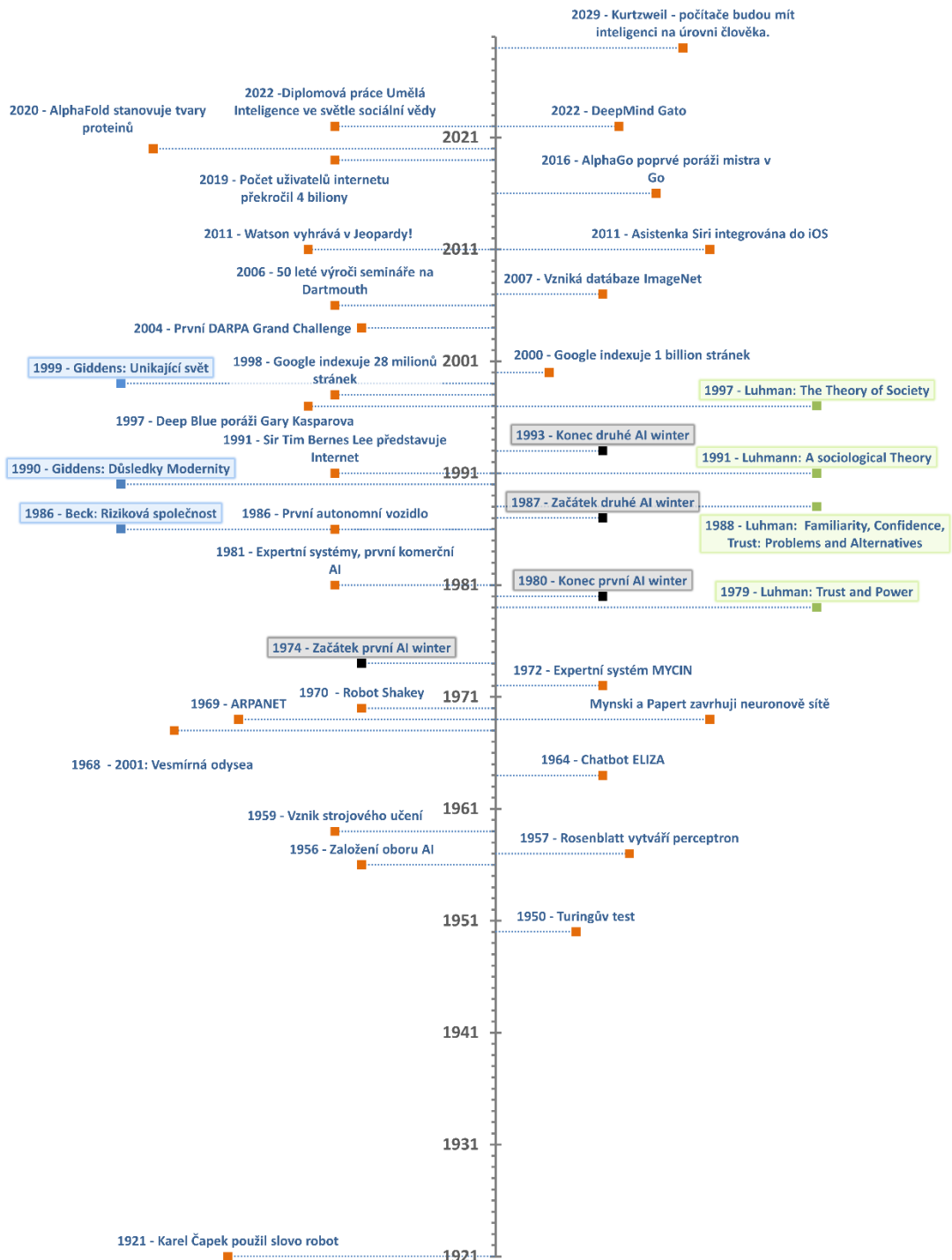
WEBER, Bruce. 1997. *Swift and Slashing, Computer Topples Kasparov*. New York Times. Dostupné z: <https://www.nytimes.com/1997/05/12/nyregion/swift-and-slashing-computer-topples-kasparov.html> 13. 6. 2022 16:16:26

WOOLDRIDGE, Michael. 2021. *A Brief History of Artificial Intelligence*. 1. vyd. Praha: Flatiron Books. 272 s. ISBN: 9781250770738

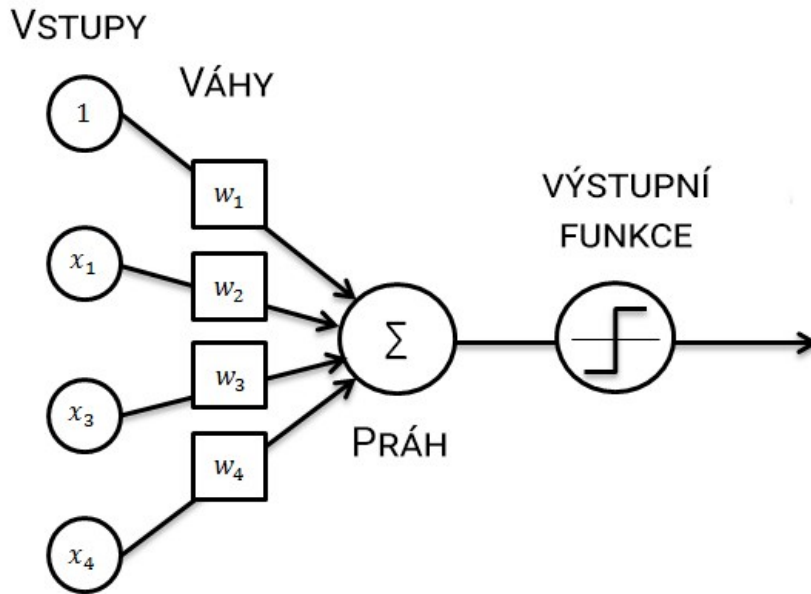
ZHANG, a kol. 2021. *The AI Index 2021 Annual Report*. Stanford Institute for Human-Centered Artificial Intelligence. Dostupné z: [https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report\\_Master.pdf](https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report_Master.pdf) 9. 6. 2022 12:42:01

## 8 Přílohy

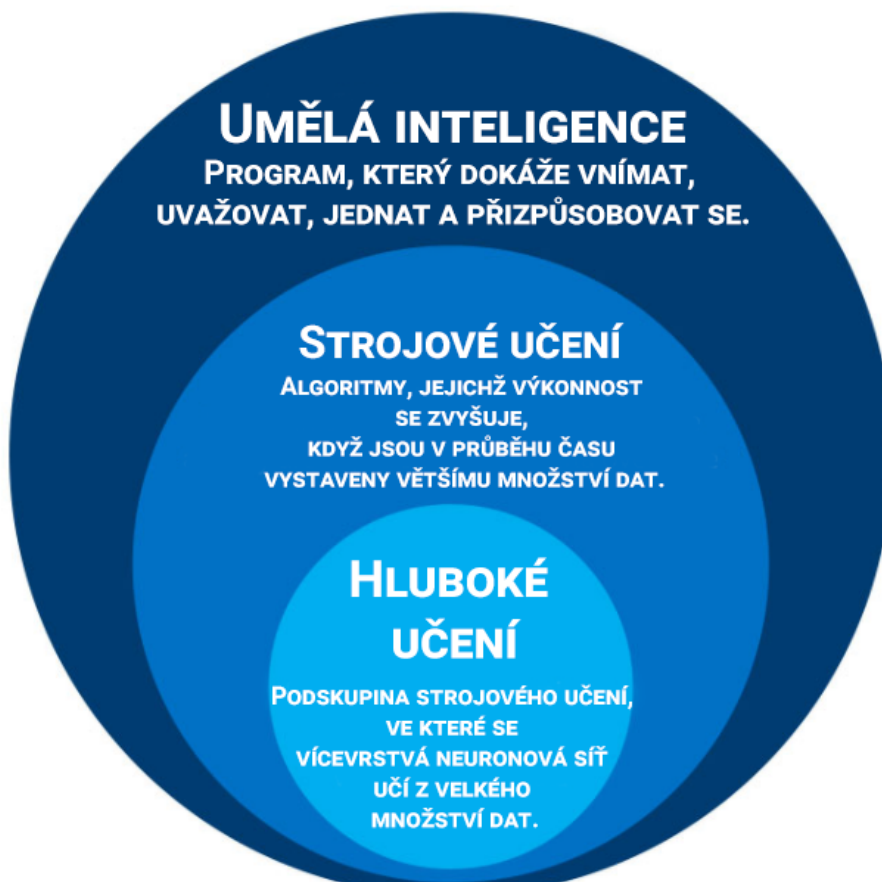
### ČASOVÁ OSA AI



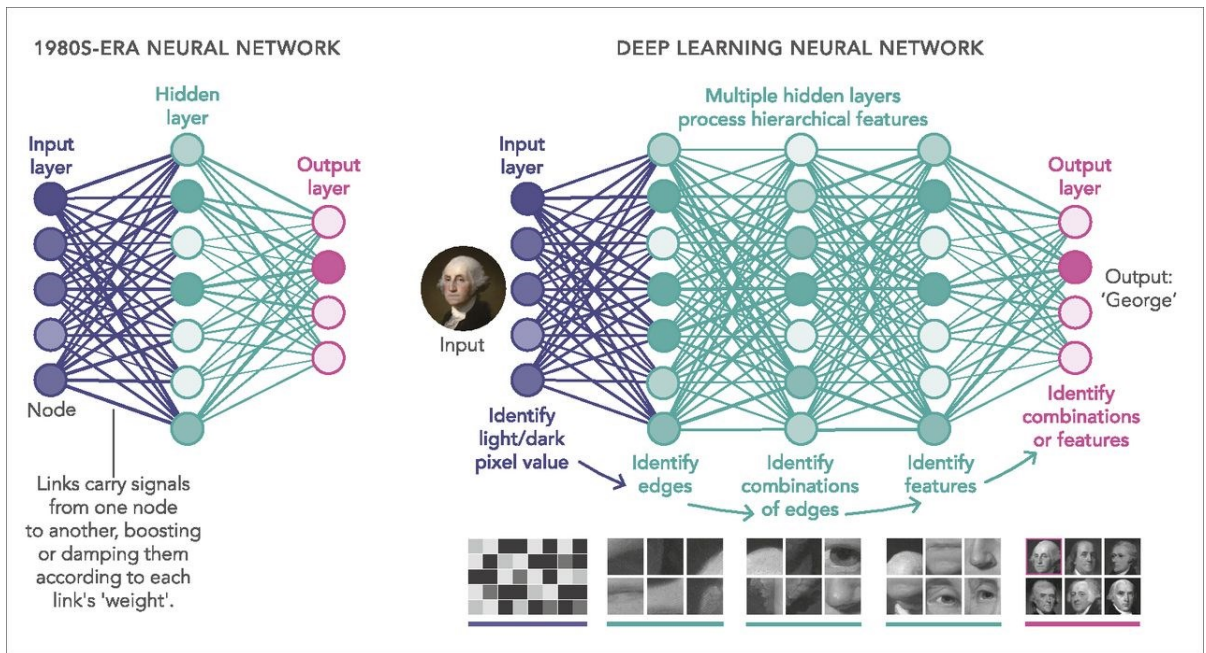
Obrázek 1 - Časová osa AI. zdroj: autor



Obrázek 2 - model perceptronu. zdroj: autor



Obrázek 3 - vztah umělé inteligence, strojového učení a hlubokého učení

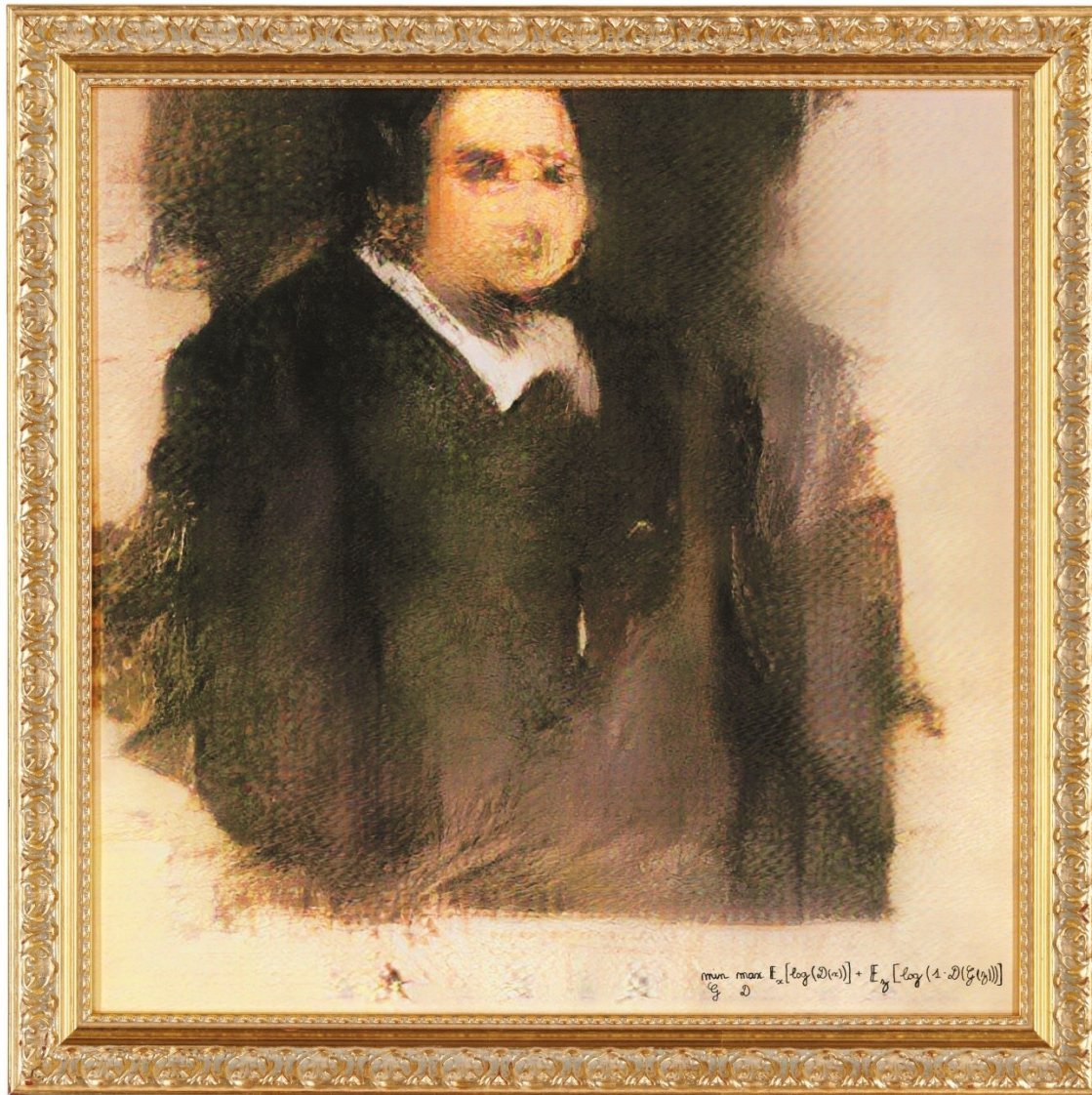


Obrázek 4 - Jednoduchá neuronová síť a hluboké učení. zdroj: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6347705/> 24. 6. 2022 22:19:34



Obrázek 5 - Autonomní vozidlo Stanley, vítěz DARPA Grand Challenge. zdroj <https://www.herox.com/blog/159-the-drive-for-autonomous-vehicles-the-darpa-grand> 24. 6. 2022 22:31:12





Obrázek 6 - První prodané dílo vytvořené AI. zdroj: <https://www.christies.com/lot/lot-edmond-de-belamy-from-la-famille-de-6166184/?from=salesummary&intobjectid=6166184&sid=18abf70b-239c-41f7-bf78-99c5a4370bc7> 28. 6. 2022 14:45:30

$$\min_G \max_D \mathbb{E}_x [\log(D(x))] + \mathbb{E}_z [\log(1 - D(G(z)))]$$

Obrázek 7 - Podpis autora prvního prodaného díla vytvořeného AI. zdroj: <https://www.christies.com/lot/lot-edmond-de-belamy-from-la-famille-de-6166184/?from=salesummary&intobjectid=6166184&sid=18abf70b-239c-41f7-bf78-99c5a4370bc7>

Napiš báseň o ochraně životního prostředí ve stylu Karla Havlíčka Borovského

Krásná je naše země,  
svou krásou nás obdarovala,  
ale my ji zničili,  
svou neznalostí a sobeckostí.

Přestože víme,  
jaká je naše chyba,  
stále ji nedokážeme napravit,  
a tak se naše země hroutí.

Pomozme jí,  
než bude pozdě,  
ať je naše země krásná i pro naše děti,  
ať již není jenom prázdnou skořápkou.

Obrázek 9 - Báseň vygenerovaná jazykovým modelem GPT-3. zdroj: autor

---

Napiš báseň o ochraně životního prostředí ve stylu Karla Hynka Máchy

Když vidím, jak krásu  
světa ničíme,  
chtělo by se mi brečet.

Vždyť to přece není tak,  
jak to má být!  
Příroda nám dala tolik,  
a my ji zahubíme.

Proč to tak děláme?  
Proč ji ničíme?  
Vždyť ona nám dává tolik,  
a my ji zahubíme.

Proč to tak děláme?  
Proč ji ničíme?  
Vždyť ona nám dává tolik,  
a my ji zahubíme.

Obrázek 8 - Báseň vygenerovaná jazykovým modelem GPT-3. zdroj: autor





AI model drawing images from any prompt!

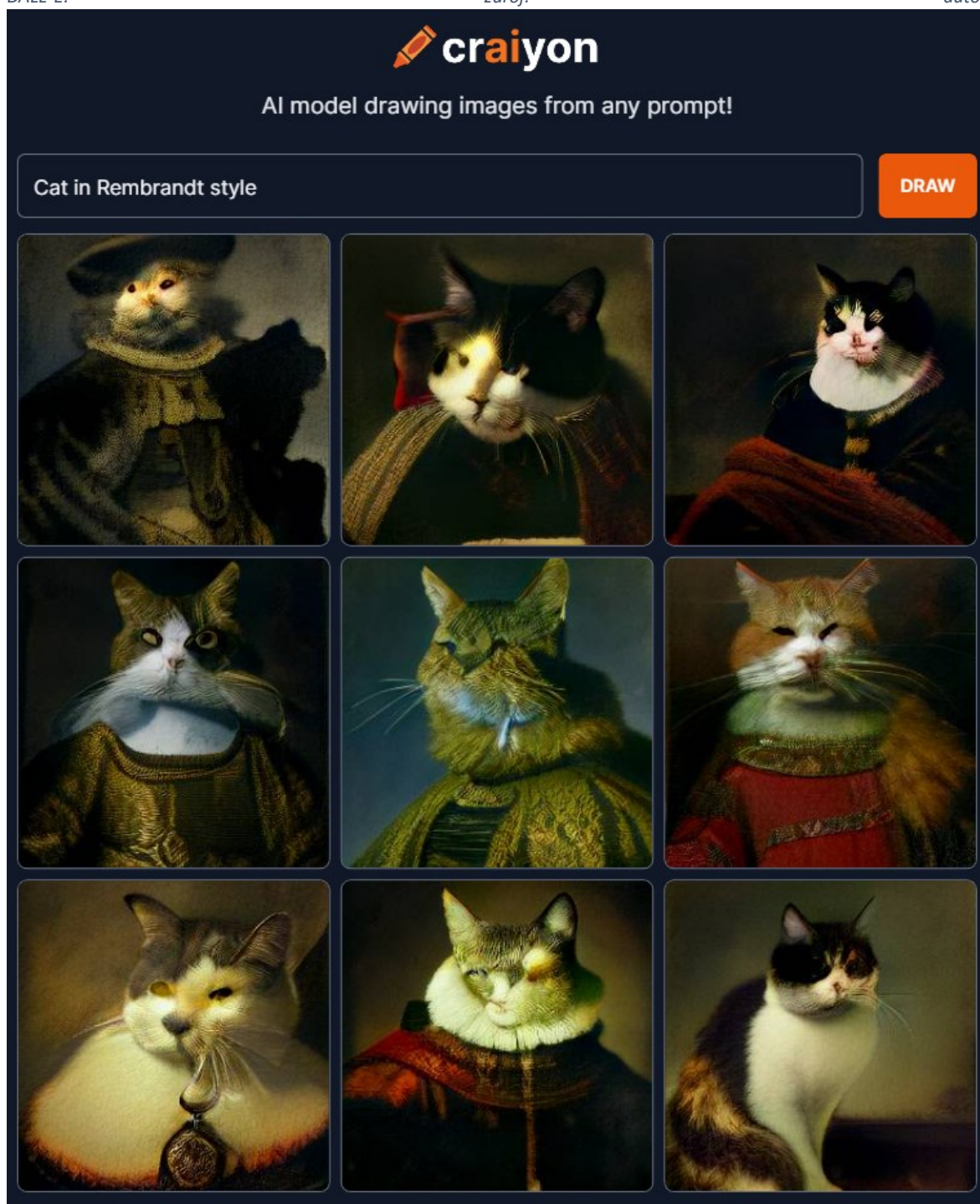
Red sports car riding into sunset on the mountain road

DRAW





Obrázek 10 - Obrázek vytvořený po zadání "Červené sportovní auto jedoucí do západu slunce na horské silnici " do modelu DALL-E. zdroj: autor



Obrázek 11 - Obrázek vytvořený po zadání "Kočka ve stylu Rembrandta" do modelu DALL-E.



AI model drawing images from any prompt!



Cat in Van Gogh style

DRAW



Obrázek 12 - Obrázek vytvořený po zadání "Kočka ve stylu Van Gogha" do modelu DALL-E.



Standard Prompting	Chain of thought prompting
<p>Example Input</p>	<p>Example Input</p>
<p>Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?</p>	<p>Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?</p>
<p>Example Output</p>	<p>Example Output</p>
<p>A: The answer is 11.</p>	<p>Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. <math>5 + 6 = 11</math>. The answer is 11.</p>
<p>Prompt</p>	<p>Prompt</p>
<p>The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?</p>	<p>The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?</p>
<p>Model Response </p>	<p>Model Response </p>
<p>The answer is 50.</p>	<p>The cafeteria had 23 apples originally. They used 20 to make lunch. So they had <math>23 - 20 = 3</math>. They bought 6 more apples, so they have <math>3 + 6 = 9</math>. The answer is 9.</p>

Obrázek 13 - Model PaLM řešící slovní úlohy včetně vysvětlení. zdroj: <https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html> 26. 6. 2022 21:34:20



**Fakulta humanitních studií UK**  
magisterský obor  
**sociální a kulturní ekologie**

Pátkova 5/2137, 182 00 Praha 8 – Libeň



Magisterský obor  
sociální a kulturní  
ekologie

**Projekt diplomové práce (DP) oboru sociální a kulturní ekologie**

1. **Jméno studenta, tituly:** Dominika Paukertová, BSc.
2. **Osobní číslo (UKČO):** 50983117
3. **Rok imatrikulace na FHS UK (bak. studium, jinak mag. studium):** 2019
4. **Datum zápisu na obor sociální a kulturní ekologie FHS UK (alespoň měsíc, rok):** září 2019
5. **Názvy všech předchozích bakalářských (magisterských) prací, škola, obor a rok, kde a kdy byly obhájeny:**  
Bakalářská práce: Business plan for Dutch e-shop  
Škola a obor: Neisse University, Information and communication management  
Obhájeno 25.6.2018, Hochschule Zittau/Görlitz, Görlitz
6. **Předběžný název DP (česky):** Umělá inteligence ve světle sociální vědy
7. **Předběžný název DP (anglicky):** Artificial intelligence in the light of social science
8. **Klíčová slova (česky):** umělá inteligence, riziko, riziková společnost, krize, společenské vědy
9. **Klíčová slova (anglicky):** artificial intelligence, risk, risk society, crisis, social sciences
10. **Obecný kontext (souvislosti tématu, širší rámec [zasazení „do světa“]):**  
Problematika umělé inteligence (dále též AI, artificial intelligence) je velice aktuální téma s dopadem na celou společnost. Ať už jste o ní slyšeli v rámci žánru sci-fi, v souvislosti s automatizací průmyslu, či jako o součásti rozličných technologií, je zřejmé, že umělá inteligence již dnes zasahuje do většiny oblastí lidského života a její vliv bude v budoucnu narůstat. O to více je zarážející, jak málo se tímto fenoménem zabývá sociální věda. Tato multidisciplinární práce si dává za cíl zkoumat umělou inteligenci jak obecně, tak i společenskovešdním prizmatem jako nový zdroj rizika.
11. **Předmět zkoumání (vlastní předmět práce [zasazení „do vědy“]):**  
Předmětem zkoumání bude analýza vybraných společenskovešdních děl a literatury zabývajících se umělou inteligencí. Pozornost bude zejména v teoretické části věnována umělé inteligenci jako takové. To se pokusím obohatit o pohled futurologů a jiných odborníků zabývajících se touto tematikou a následně reflektovat ve vybraných dílech sociální vědy. To bude realizováno v rámci analytické části, která nabídne hledání podobností a rozdílů v obou typech literatury, zejména hledání toho, na jaké podoby, výhody a nevýhody, příp. rizika AI jsou společenskovešdní disciplíny schopné upozornit a kdy (v jakém stádiu vývoje AI). Tato analýza nám napoví, zda je (v případě technologií někdy opomíjená) společenská věda schopná přinést relevantní poznání i v technickém oboru a případně i předvídat nebo předjímat budoucnost.

12. Hlavní vstupní výzkumný problém – výzkumná otázka (výzkumné otázky) – ev. hypotéza (hypotézy):

- Dokáží společenské vědy anticipovat budoucnost v případě umělé inteligence?
  - Kde se ne/objevují jednotlivé pojmy v různých typech literatury?
  - Jak se liší pojetí rizika v jednotlivých typech literatury?
  - V jakých kontextech se jednotlivé pojmy objevují?
  - S jakými důsledky se jednotlivé pojmy vyskytují?
  - Jsou témata, která jeden nebo druhý typ literatury nepopisuje?
  - Jak se liší jazyky a obsahy obou typů popisu (v hlavních rysech)?

13. Metodologický postup: metody a techniky, které budou v práci použity:

Prvním krokem bude literární rešerše vybrané společenskovední literatury, zejména té, jež tematizuje riziko, a literatury zabývající se umělou inteligencí a z ní si vymezím několik pojmů a témat, kterým se budu věnovat. V dalším kroku budu za pomoci Althusserovy metody symptomálního čtení a hermeneutických postupů zkoumat a srovnávat, jak vybraná díla dané pojmy a témata zobrazují a uchopují.

14. Cíl DP (kromě ověření hypotéz a teoretického přínosu např. praktický přínos, vypracování metodologie, základ pro řešení problémů v praxi atd.):

Spojení umělé inteligence a společenské vědy je stále ještě málo objevené téma, zvláště pak na české akademické půdě. Tato práce by mohla obohatit vědění týkající se umělé inteligence o společenskovední pohled a navodit zájem o tuto oblast.

15. Čím budou rozšířeny dosavadní znalosti (vědecká „přidaná hodnota DP“):

Jak již bylo řečeno výše, umělá inteligence je velice relevantní, leč ve společenských vědách málo reflektované téma. Tato práce se snaží o alespoň o částečné rozpracování tohoto fenoménu. Jedním z hlavních přínosů by mělo být zjištění, zda, jak a v jakém předstihu dokáží společenskovední disciplíny předjímat podobu, uplatnění a dopady AI.

16. Jaké bude (bude-li) jejich teoretické zobecnění a přínos:

Pravděpodobně žádné.

17. Struktura DP (předběžný obsah – názvy oddílů a kapitol):

- Úvod
- Teoretická část
  - Umělá inteligence
  - Riziko
  - Úvod k analýze
    - Vybrané pojmy
    - Vybraná témata
- Analytická část
  - Typy zdrojů k analýze
  - Srovnání

- Pojmy
- Témata
- Kontexty
- Odlišnosti

- Diskuze
- Závěr

18. Předběžná bibliografie k tématu:

BECK, Ulrich. 2018. *Riziková společnost: Na cestě k jiné modernitě*. 3. vyd. Praha: Sociologické nakladatelství. 431 s. ISBN: 978-80-7419-267-8

BOSTROM, Nick. 2017. *Superintelligence: Superintelligence: Až budou stroje chytřejší než lidé*. 1. vyd. Praha: Prostor. 512 s. ISBN 978-80-7260-353-4

BRYNJOLFSSON, Erik; MCAFEE, Andrew. 2015. *Druhý věk strojů*. 1. vyd. Jan Melvil Publishing. 296 s. ISBN: 978-80-87270-71-4

FLORIDI, Luciano. 2014. *Fourth Revolution*. 1. vyd. UK: Oxford University Press. 272 s. ISBN: 0199606722

FORD, Martin. 2016. *Rise of the Robots*. 1. vyd. USA: Basic Books. 368 s. ISBN: 9780465097531

GASSET, José Ortega y. 2011. *Úvaha o technice*. (Úvaha o technice a jiné eseje o vědě a filosofii.) Přel. Špína, Michal. 1. vyd. Praha: Oikoymenth. 127 s. ISBN 978-80-7298-455-8.

GIDDENS, Anthony. 2000. *Unikající svět (jak globalizace mění náš život)*. 1. vyd. Praha: Sociologické nakladatelství (SLON). 135 s. ISBN 80-85850-91-5.

GIDDENS, Anthony. 2010. *Důsledky modernity*. 3. vyd. Praha: Sociologické nakladatelství. 158 s. ISBN: 978-80-7419-035-3

GRONDIN, Jean. 1997. *Úvod do hermeneutiky*. Z něm. orig. přel. Horyna, Břetislav; Kouba, Pavel. 1. vyd. Praha: Oikoymenth. 247 s. ISBN 80-86005-43-7

HAUSER, Michael. *Althusserovo symptomální čtení jako lekce historického materialismu*. dostupné z: <https://sok.bz/clanky/2011/michael-hauser-althusserovo-symptomalni-cteni-jako-lekce-historickeho-materialismu> 11.5.2021 17:58

HENDL, Jan. 2005. *Kvalitativní výzkum. Základní teorie, metody a aplikace*. 2. vyd. Praha: Portál. 408 s. ISBN 978-80-7367-485-4

KUŽEL, Petr. *Filosofie Louise Althussera: o filosofii, která chtěla změnit svět*. Praha: Filosofía, 2014. ISBN 9788070074176

ORD, Toby. 2020. *The Precipice: Existential Risk and the Future of Humanity*. 1. vyd. UK: Bloomsbury. 480 s. ISBN: 1526600226

PERROW, Charles. 2010. *Normal accidents: Living with High-Risk Technologies*. 3. vyd. USA: BasicBooks. 366 s. ISBN: 0-465-05142-1

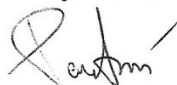
POKORNÝ, Petr a kol. 2005. *Hermeneutika jako teorie*. Praha: Vyšehrad. ISBN: 80-7021-779-0

POKORNÝ, Petr; STORCH, David a kol. 2020. *Antropocén*. 1. vyd. Praha: Academia. 660 s. ISBN: 978-80-200-3129-7


19. Předpokládaný vedoucí DP: PhDr. Ivan Rynda

20. Důvod volby tématu (dosavadní znalosti, zájem, praxe a zájem studenta):<sup>1</sup>

Troja 17. května 2021

  
diplomant/ka

  
vedoucí DP

  
zást. garanta SKE pro studijní záležitosti

---

<sup>1</sup> nepovinné