

**FILOZOFICKÁ FAKULTA UNIVERZITY PALACKÉHO  
V OLOMOUCI**

**KATEDRA SLAVISTIKY**

**Komentovaný překlad odborného textu  
z technologické oblasti: Sociální a mezinárodní  
dopady využití technologie umělé inteligence**

**Annotated translation of a specialized text from technological  
field: Social and international impacts of using artificial  
intelligence technology.**

**Комментированный перевод специального текста из  
технологической сферы: Социальные и международные  
последствия использования технологии искусственного  
интеллекта**

**VYPRACOVAL: Bc. Alexandr Oros**

**VEDOUCÍ PRÁCE: Mgr. Martina Pálušová, Ph.D.**

**2024**

Prohlašuji, že jsem práci vypracoval samostatně a uvedl všechny použité prameny a literaturu.

V Olomouci, 9. 5. 2025

---

Podpis

Děkuji Mgr. Martině Pálušové, Ph.D., za konzultace, rady a připomínky, které mi během psaní magisterské práce poskytla.

V Olomouci, 9. 5. 2024

---

Podpis

# Obsah

<b>ÚVOD</b> .....	<b>1</b>
<b>1. TEORETICKÁ ČÁST</b> .....	<b>2</b>
1.1 ÚVOD DO TEORIE .....	2
1.2 STRUČNÁ HISTORIE UMĚLÉ INTELIGENCE V PŘEKLADU .....	2
1.3 SLEDOVÁNÍ PŘEKLADATELŮ BĚHEM PRÁCE .....	5
1.4 STROJOVÝ PŘEKLAD V LÉKAŘSTVÍ A PRÁVU .....	6
1.3.1. <i>Strojový překlad v lékařství</i> .....	7
1.3.2. <i>Strojový překlad v právu</i> .....	9
1.5 ETIKA A UMĚLÁ INTELIGENCE .....	10
1.4.1. <i>Vlastnictví dat a jejich využití</i> .....	11
1.4.2. <i>Osobní údaje a soukromí</i> .....	12
1.4.3. <i>Předsudky vůči pohlaví ve strojovém překladu</i> .....	14
1.6 HODNOCENÍ A POST-EDITACE STROJOVÉHO PŘEKLADU .....	16
<b>2. PRAKTICKÁ ČÁST</b> .....	<b>20</b>
2.1. CHARAKTERISTIKA PŘEKLÁDANÉHO TEXTU .....	20
2.2. ANALÝZA PŘEKLÁDANÉHO TEXTU .....	20
2.2.1. <i>Kompozice textu</i> .....	20
2.2.2. <i>Syntaktická analýza</i> .....	21
2.2.3. <i>Morfologická analýza</i> .....	23
2.2.4. <i>Lexikální analýza</i> .....	24
2.2.5. <i>Obsahová analýza</i> .....	25
<b>3. PRAKTICKÁ ČÁST – PŘEKLAD</b> .....	<b>26</b>
<b>4. PŘEKLADOVÝ KOMENTÁŘ</b> .....	<b>49</b>
4.1. TRANSLITERACE .....	49
4.2. TRANSFORMACE SLOVOSLEDU .....	50
4.3. UNIVERBIZACE .....	51
4.4. GENERALIZACE .....	51
4.5. MULTIVERBIZACE .....	52
4.6. KALKOVÁNÍ .....	52
4.7. ZÁMĚNA GRAMATICKÉ KATEGORIE .....	52
<b>ZÁVĚR</b> .....	<b>54</b>
<b>PEŽIOME</b> .....	<b>55</b>
<b>ANOTACE</b> .....	<b>63</b>



<b>BIBLIOGRAFIE .....</b>	<b>64</b>
<b>PŘÍLOHA Č. 1 – ORIGINÁLNÍ TEXT .....</b>	<b>68</b>
<b>PŘÍLOHA Č. 2 – TERMINOLOGICKÝ GLOSÁŘ.....</b>	<b>96</b>

# Úvod

Předmětem mé magisterské práce je komentovaný překlad odborného textu s komentářem a glosářem. Překládaný text pochází z technologické oblasti, s konkrétním zaměřením na technologii umělé inteligence (UI), který zkoumá její mezinárodní a společenské dopady.

Toto téma jsem si vybral kvůli jeho aktuálnosti a celosvětovému významu. Umělá inteligence má významný vliv na mnoho oblastí života, včetně průmyslu, technologií a komunikace. Její rostoucí význam a dopad na mezinárodní vztahy a společnost mě přivedly k volbě tohoto tématu, které mi poskytne přehled o UI a přispěje k mé budoucí překladatelské praxi. Zároveň je téma umělé inteligence v drtivé většině případů prezentováno pozitivně, ale jeho potenciální negativní dopady jsou zřídka rozebírány.

Cílem této práce je v první řadě provést přesný a ekvivalentní překlad textu o umělé inteligenci a zároveň analyzovat jeho obsah. Dalším cílem je poskytnout rozbor jazykových a lingvistických termínů, včetně odborné terminologie, stylistických prvků a lexikálních obrátů. V neposlední řadě je mým cílem ukázat, jak UI ovlivňuje různé oblasti.

Předkládaná práce se skládá ze dvou částí, a to teoretické a praktické. V teoretické části se nejprve věnuji historii a vývoji umělé inteligence a jejímu vlivu na strojový překlad, včetně etických otázek, jako je ochrana osobních údajů, vlastnictví dat a genderové předsudky. Také se budu zabývat využíváním strojového překladu v oblastech, jako jsou medicína a právo. Tyto oblasti vyžadují maximální přesnost překladu, jelikož i sebemenší chyba může vést k fatálním následkům. Využívání strojového překladu se nicméně v těchto oblastech stává čím dál běžnějším, s čímž i narůstá riziko pochybení, které může vést ke špatným soudním rozhodnutím, zdravotním následkům či dokonce úmrtí. V krátkosti se také zaměřím na post-editaci strojového překladu, jakožto na narůstající způsob úpravy textu. Kvalita překladů za pomoci umělé inteligence je natolik vysoká, že potřeba mechanického překladu klesá, zatímco potřeba post-editace textů, přeložených umělou inteligencí, naopak narůstá. Následuje praktická část, která bude začínat rozбором překládaného textu a jeho analýzou z hlediska kompozice, syntaktických, morfologických a lexikálních prvků a odborné terminologie.

Za obsahovým a formálním rozбором bude následovat samotný překlad odborného textu o UI a rozbor nejčastěji využívaných jazykových transformací při převodu výchozího textu do cílového jazyka. Součástí práce jsou také dvě přílohy: terminologický glosář a originální text.

# 1 Teoretická část

## 1.1 Úvod do teorie

Umělá inteligence stále více ovlivňuje různé oblasti života, včetně strojového překladu, a má výrazné dopady na životy spousta lidí po celém světě. I přes to, že je tento fenomén veřejně prezentován především z pozitivního hlediska, a bezpochyby přináší mnoho výhod a ulehčuje každodenní rutinní činnosti, má také spoustu negativních dopadů, o kterých se zřídka mluví.

Jelikož se překládaný text v praktické části práce zabývá především společenskými a mezinárodními dopady UI, zejména v kontextu jejího vlivu na pracovní trh, mezinárodní vztahy a technologickou suverenitu, a popisuje, jak UI transformuje různé průmyslové odvětví, a ovlivňuje mezinárodní bezpečnost a stabilitu, budu se v teoretické části věnovat využití UI ve strojovém překladu, kde se zaměřím na její roli v překonávání jazykových bariér a jejím použití v různých oblastech, jako je lékařství a právo.

Budu se také zabývat etickými otázkami, jako je ochrana osobních údajů, vlastnictví dat a předsudky, které mohou ovlivnit výstupy strojového překladu. Tato část tedy poukazuje na to, jak může UI usnadnit komunikaci, ale zároveň přinést zdravotní, právní, etická a bezpečnostní rizika spojená s jejím využitím. Zdůrazním také důležitost regulace a zodpovědného využití, které mohou zmírnit tyto negativní, potenciálně fatální důsledky.

## 1.2 Stručná historie umělé inteligence v překladu

Pokusy o automatizaci překladů sahají až do 17. století, ale reálné možnosti se objevily teprve ve 20. století. V 30. letech 20. století podali Georges Artsrouni a Petr Trojanskij žádosti o patenty na překladové stroje. Trojanskij navrhl nejen automatický slovník, ale také systém kódování gramatických funkcí mezi jazyky. Jeho myšlenky však zůstaly neznámé až do konce 50. let, kdy se již objevily první počítače. (Hutchins 2014)

Strojový překlad prošel značným vývojem v 50. letech 20. století. Počáteční modely strojového překladu, založené na jednoduchých systémech založených na pravidlech (rule-based), se postupně vyvíjely až k dnešním pokročilým neuronovým sítím, které se staly standardem ve většině aplikací strojového překladu. Tento vývoj je znatelný zejména v posledních letech, kdy byla kvalita překladů výrazně posílena díky začlenění technik hlubokého učení.

Výzkum strojového překladu začal krátce po vynálezu prvního počítače a je považován za jednu z největších výzev v oblasti zpracování přirozeného jazyka. Warren Weaver představil koncept strojového překladu již v roce 1947, což ukazuje na dlouhodobý zájem o automatický překlad mezi jazyky.

Původní strojový překlad byl založen na pravidlech (rule-based), při kterém systémy využívaly dvojjazyčné slovníky a ručně psaná pravidla k překladu textů. Překlad založený na pravidlech pracuje s morfologií, syntaxí a sémantikou obou jazyků. Vyžaduje tedy syntaktickou a sémantickou analýzu zdrojového textu, a pro vytvoření překladu v cílovém jazyce je potřeba generování syntaxe a sémantiky (Irfan 2017). Tento přístup však byl pracný a pravidla byla obtížně udržitelná a nepřenositelná mezi různými jazyky a oblastmi, což omezovalo jejich škálovatelnost. Přesto byl strojový překlad zpočátku využíván především pro vojenské účely a v roce 1954 byl na Georgetovské univerzitě proveden první rusko-anglický experiment s využitím počítače IBM. Strojový překlad přitáhl velkou pozornost až do vydání zprávy ALPAC v roce 1966, která výzkum strojového překladu velmi zkritizovala, což vedlo k výraznému snížení financování.

S nástupem dostupnosti dvojjazyčných korpusů se v novém tisíciletí začaly více uplatňovat takzvané „corpus-based“ metody, zahrnující „example-based machine translation“ (EBMT), „statistical machine translation“ (SMT) a „neural machine translation“ (NMT). EBMT byl navržen v polovině 80. let a zaměřoval se na vyhledávání podobných větných párů ve dvojjazyčném korpusu. Přestože přinášel kvalitní výsledky, pokrýval jen omezený rozsah překladů, což vedlo k jeho využívání hlavně v CAT (computer-aided translation) systémech (Wang 2021:143).

SMT, navržený Brownem a spol. v roce 1990, umožnilo automatické učení překladových pravidel z velkého množství dat bez potřeby jejich ručního psaní. Tato metoda byla formalizována v roce 1993 a i přes počáteční složitost a převahu RBMT v komerčním prostředí začala postupem času narůstat na popularitě. SMT bylo podpořeno vydáním nástrojů jako Egypt, GIZA a GIZA++, a následně se rozšířily i SMT metody založené na frázích, které dále zlepšily kvalitu překladů. V roce 2006 Google zavedl svůj překladač založený na frázové SMT metodě.

Statistický strojový překlad (SMT) považuje překlad za problém strojového učení. To znamená, že pomocí učebního algoritmu pracujeme s velkým množstvím dříve přeloženého textu, který se označuje jako paralelní korpus, paralelní text, bitext nebo multitext. (Lopez 2008)

Vývoj v oblasti strojového překladu (MT) a technik zpracování řeči vedl k pokrokům v simultánním tlumočení. První pokusy o automatický překlad řeči sahají až do roku 1983, kdy byl představen experimentální tlumočnický systém. Od té doby došlo k nárůstu a zlepšení systémů pro simultánní překlad. S příchodem neurálních strojových překladů a neurálního rozpoznání řeči se tyto nové systémy snaží zautomatizovat proces tlumočení, který je pro člověka mimořádně náročný. Rozvoj metod pro simultánní tlumočení by mohl významně ulehčit práci tlumočnickům a zároveň zpřístupnit tlumočnické služby širší veřejnosti za nižší ceny. V posledních letech se v oblasti simultánního tlumočení angažovaly přední výzkumné týmy a byly spuštěny komerční produkty, které jsou využívány na mnoha mezinárodních konferencích, čímž došlo k obnovenému zájmu o tuto technologii.

V poslední době došlo k významnému pokroku v oblasti neurálního strojového překladu (NMT). Typický model NMT se skládá ze dvou částí: kódovací síť, která převede větu ze zdrojového jazyka na reálné hodnoty, a dekodovací síť, která na základě tohoto vektoru vytvoří překlad. Tento strojový překlad je založen na principu neurální sítě, což je metoda strojového učení, která zpracovává různé vstupy a předpovídá výstupy (Koehn 2017: 6). Tento proces lze přirovnat k lidskému překladu, kdy model nejdříve "přečte" celou větu ve zdrojovém jazyce a pak na základě svého porozumění této větě vygeneruje překlad slovo po slově. Na rozdíl od předchozích metod jako je RBMT (rule-based machine translation) a SMT (statistical machine translation), NMT nepotřebuje pravidla navržené člověkem. NMT je komplexní systém, který se přímo učí sémantické reprezentace a znalosti překladu z trénovacích korpusů. Díky těmto výhodám se NMT stalo dominantní metodou ve strojovém překladu.

Neurální strojové překlady (NMT) zaznamenaly značné zlepšení ve schopnosti strojů zpracovávat kontext a nepatrné jazykové rozdíly, což dramaticky zvýšilo kvalitu překladů a efektivitu procesu překladu. Děti se učí jazyk tím, že poslouchají ostatní a rozpoznávají struktury a vzorce v jazyce. Ve světě umělé inteligence se tomu říká neurální strojový překlad. Tento pokrok umožnil strojovému překladu stát se důležitým nástrojem v mnoha aplikacích, od automatického překladu webových stránek po aplikace pro překlad v reálném čase (Liu, Afzaal 2021: 66).

Přestože byl v oblasti strojového překladu učiněn značný pokrok, stále existují výzvy, jako je překlad mezi jazyky s odlišnými strukturami a zachování kulturních rozdílů. Tyto problémy poukazují na omezení současných technologií a na oblasti, které vyžadují další výzkum (Wang 2021:143, 144, 150).

### 1.3 Sledování překladatelů během práce

Vzhledem k rostoucí globální poptávce po překladech čelí překladatelé, dodavatelé i zákazníci zvýšenému tlaku v důsledku větších objemů práce a těsnějších termínů. V tomto kontextu hraje klíčovou roli efektivita a produktivita, které umožňují včasné dodání. Prudký technologický rozvoj spojený s příchodem umělé inteligence vedl k rozšíření metod hodnocení překladových produktů a procesů, které jsou nyní snadno dostupné. Sledování činnosti je jednou z takových metod, která je používána nejen v překladatelském průmyslu, ale i v mnoha dalších oblastech. Evropská unie například financovala projekty, které zahrnovaly sledování v oblastech zdraví, bezpečnosti, elektřiny a mobility. Diskuse o sledování mají tedy širší význam v kontextu umělé inteligence a dohledu a staly se součástí našich životů od pandemie Covid-19.

V kontextu překladu zahrnuje sledování činnosti sběr různých behaviorálních dat od překladatelů během práce na textu, například zaznamenáním času a aktivit na klávesnici. Mnoho překladatelských společností již používá nějakou formu sledování k hodnocení produktivity svých překladatelů. Některé společnosti nyní sledují i externí překladatele, ať už jde o výpočet hodinové mzdy nebo monitorování pokroku. I když může být sledování v určitých případech oprávněné, praxe je kontroverzní a v zásadě neregulovaná. Některé online překladatelské platformy již implementují sledování jako výchozí možnost. Nicméně sledování by mohlo mít také potenciálně pozitivní aplikace, například jako prostředek pro překladatele, aby si uvědomili své pracovní procesy, nebo k zaručení práv interních překladatelů, tj. aby byly dodrženy jejich přestávky a pracovní kalendář. Přestože jsou tyto perspektivy protichůdné, překvapivě málo výzkumů se zabývalo tím, jak je sledování činnosti vnímáno a jak jsou nástroje pro zvyšování produktivity používány profesionálními překladateli v jejich každodenní práci.

Studie „Translation Productivity“ se zaměřila na profily různých účastníků, nástroje pro sledování a proměnné produktivity, následované procesy sběru dat projektu a dotazníku. Tento výzkum odhalil různorodé pocity účastníků v průběhu studie. Překladatelé byli požádáni, aby popisovali své pocity vůči sledování své aktivity a používání nástrojů pro zvýšení produktivity. Zajímavým zjištěním z této části dotazníku bylo kolísání osobních pocitů, jak studie postupovala.

Interní překladatel obecně vyjadřoval velmi pozitivní názory na význam měření vlastní produktivity a zřídka kritizoval nástroje pro sledování. Přesto při popisování svých pocitů

vyjádřil protichůdný postoj k sledování, naznačující, že někdy to může působit jako břemeno a vyvolává pocit, že je nutné neustále překračovat určitý práh. Externí překladatel zažíval naopak během studie kolísání pocitů: zpočátku byl "zvědavý", když byla zkušenost čerstvá, nicméně s časem se začaly objevovat etické úvahy o tom, jak by nástroje pro sledování mohly být klienty využívány, a sledování se mohlo zdát "vtíravé", "nevhodné" a "narušující". Externí překladatel naznačil, jak by sledování produktivity mohlo vést k profilování a srovnávání.

Externí překladatel také poznamenal, jak může být sledování používáno jako forma posuzování. Většina překladatelů v této studii si byla vědoma potenciálních problémů, které přináší technologie sledování. Na rozdíl od obvyklého předpokladu, že vyšší produktivita znamená rychlejší práci a že rychlost je vždy dobrá, poukázali překladatelé na to, že 'být pomalý' může být pozitivní, pokud je to prezentováno jako znak pečlivosti. Překladatelé také zdůraznili důležitost autonomie pro profesionální externí překladatele, sledování by tak mohlo být zvláště problematické pro překladatele samostatně výdělečně činné.

Výzkum zdůrazňuje nezbytnost zavedení mechanismů souhlasu při práci online a na zvýšení transparentnosti platform ohledně sbíraných dat a důvodů jejich sběru. Zainteresované strany by měly spolupracovat na implementaci etických praxí.

Jelikož může sledování podporovat nespravedlivé praktiky vůči překladatelům a regulace sledování aktivit zatím de facto neexistují, je důležité další zkoumání právních důsledků této problematiky.

Data naznačují, že praktiky sledování pravděpodobně ovlivňují externí a interní překladatele různě. Negativní následky sledování aktivit mohou být pro samostatně výdělečně činné překladatele vnímány intenzivněji, jelikož si mnozí z nich zvolili práci na volné noze právě kvůli vyhnutí se typu monitorování a kontroly, které sledování aktivit obvykle zahrnuje (Ragni et. al., 2022: 4, 20-25).

#### **1.4 Strojový překlad v lékařství a právu**

Využití strojového překladu se stalo běžnou součástí překonávání jazykových bariér nejen v osobním, ale i veřejném životě. Jeho pronikání do rizikových oblastí, jako jsou zdravotnická zařízení a soudní síně, s sebou nese důležité otázky bezpečnosti a přesnosti. Přestože technologie strojového překladu nepřestává postupovat kupředu, a v současné

době je reprezentována především neurálními sítěmi, které produkují plynulé a zdánlivě přirozené překlady, musíme si být vědomi i možných rizik plynoucích z jejich používání. Nebezpečí chybně přeložených frází, které mohou mít v medicínském kontextu fatální důsledky, nesprávně interpretovaných právních termínů mohou naopak vést k justičním omylům, což vyžaduje důkladnější zkoumání a interdisciplinární přístup k hodnocení dopadů strojového překladu.

Studie poukazují na to, že i v pokročilých systémech, jako je Google Translate, který je používán k překladu miliard vět ročně, existují značné slabiny, zvláště když se jedná o překlad neobvyklých nebo kriticky důležitých frází. Tento stav zvyšuje potřebu lepšího porozumění limitům technologie a současně zdůrazňuje důležitost většího povědomí o kapacitách i omezeních strojového překladu u široké veřejnosti. V této souvislosti je klíčové rozvíjet výzkum a strategie, které by napomohly bezpečnějšímu a informovanějšímu používání strojový překlad v oblastech, kde je riziko chyb kriticky vysoké (Vieira, O'Hagan, O'Sullivan 2020:1515).

### **1.3.1. Strojový překlad v lékařství**

Jan Hendl rozděluje překlad lékařského jazyka na dvě klíčové oblasti. První se zabývá překladem lékařské terminologie mezi různými jazyky, zatímco druhá oblast se zaměřuje na převod odborného lékařského dialogu do výrazů srozumitelných širší veřejnosti. Obě oblasti jsou důležité, protože zlepšují komunikaci, ať už mezi jazyky, nebo mezi různými skupinami lidí. (Hendl 2023: 8)

Překlad lékařských pojmů mezi jazyky může zjednodušit mezinárodní spolupráci ve výzkumu i při zavádění lékařských technik. Jazyk je často jednou z hlavních překážek celosvětové spolupráce, a s pomocí velkého jazykového modelu (LLM) lze tuto bariéru výrazně snížit.

Pacienti s omezenou znalostí anglického jazyka čelí komunikačním bariérám ve zdravotnictví anglicky mluvících zemí. Přestože písemná komunikace zvyšuje porozumění, standardně předložené instrukce nemohou reflektovat specifika pacienta, například v otázkách dávkování léků.

Při klinickém vyšetření může lékař, který se potýká s jazykovou bariérou a nemá k dispozici profesionálního tlumočnicka, použít strojový překladatel ke komunikaci s pacientem. Strojový překlad lze využít k upřesnění anamnézy pacienta, přezkoumání



klinické diagnózy nebo objasnění doporučeného plánu léčby a následné péče (Randhawa et. al. 2013). Nástroje pro strojový překlad, včetně Google Translate (GT), mají potenciál zlepšit komunikaci s těmito pacienty, nicméně dřívější studie ukázaly jejich omezenou přesnost. Jedna studie odhalila, že španělské překlady informačních materiálů pro pacienty od GT dosahovaly 60% přesnosti, přičemž 4 % překladů obsahovalo závažné chyby. V roce 2017 Google Translate aktualizoval svůj algoritmus překladu a prohlašoval o jeho významném zlepšení. V této studii byla hodnocena schopnost GT překládat pokyny k propuštění z pohotovosti do španělštiny a čínštiny.

Přestože nový algoritmus GT přinesl vyšší přesnost a menší počet chyb s vážnými následky ve srovnání s dřívějšími výsledky, bylo zjištěno, že 2 % překladů do španělštiny a 8 % překladů do čínštiny mělo potenciál způsobit významné zdravotní riziko. GT tedy může sloužit jako doplňková podpora (nikoli náhrada) písemných anglických instrukcí, ale měl by obsahovat varování o možné nepřesnosti překladů (Khoong, Elaine C. et al 2019).

Také bylo provedeno vyhodnocení deseti běžně používaných lékařských frází ve 26 jazycích, což zahrnovalo jazyky západní a východní Evropy, Asie a Afriky. Z 260 překladů bylo 150 překladů správně, nicméně téměř polovina (110) obsahovala vážné chyby. Nejnižší úspěšnost měly africké jazyky, za nimi následovaly asijské jazyky, zatímco západní evropské jazyky dosahovaly nejlepších výsledků. Nejlépe přeloženou frází mezi všemi jazyky byla fráze „*Your husband has the opportunity to donate his organs*“, neboli: „Váš manžel má možnost darovat orgány“, zatímco fráze „*Your child has been fitting*“, neboli „Vaše dítě má záchvaty“, bylo přeloženo správně pouze v necelých osmi procentech případů. V některých jazycích došlo k vážným překladovým chybám, jako například překlad do svahilštiny, kde bylo dítě označeno za mrtvé, místo toho, aby bylo uvedeno, že dostává záchvaty. Další významné chyby byly zaznamenány také v polštině, maráthštině a bengálštině, kde došlo k nesprávným a často až komickým interpretacím lékařských termínů (Patil a Davies 2014).

V dalším příkladu byly anglické články o depresivních poruchách přeloženy do zjednodušené čínštiny pomocí Google Translate. Přijatelnost sémantického rozdílu mezi původními anglickými texty a jejich automaticky přeloženými verzemi byly posuzovány po zpětném překladu do angličtiny. Na základě zpětných překladů byly odhaleny chyby symptomů depresí.

V následující tabulce jsou ilustrativní příklady chyb překladu způsobených nástrojem NMT (neurální strojový překlad). Při použití Google Translate bylo také zjištěno, že slovo „deprese“ bylo konzistentně přeloženo jako „frustrace“ do čínštiny,

slovo „vina“ bylo přeloženo jako „ostuda“ do hindštiny a slovo „soudný“ bylo přeloženo jako „kritický“ do španělštiny. Pro uživatele bez dvojjazyčných dovedností, jako jsou profesionální překladatelé k ověření pravosti automatického překladu, nebo bez zdravotnické gramotnosti, jako jsou lékaři, může být použití neurálního strojového překladu (NMT) k posouzení nebo diagnostice duševních zdravotních stavů zavádějící, což může vést k diskriminačním postojům nebo chování vůči jednotlivcům projevujícím mírné až těžké symptomy duševní poruchy (Want et. al., 2021).

**Obr. č. 2** (Zpětný překlad do angličtiny z čínštiny pomocí Google Translate)

	Original english material	Backtranslation from simplified chinese using GT
1	A low mood that doesn't lift may be a sign of depression. However, depression isn't just about low mood. You may experience the following: Feeling bored or <b><u>restless</u></b> .	Low mood that doesn't get better may be a sign of depression. However, depression is more than just a low mood. You may encounter the following situations: Feeling bored or <b><u>upset</u></b> .
2	Harassment is behaviour which frightens you and causes you <b><u>distress or alarm</u></b> .	Harassment is behaviour that scares you and causes you <b><u>pain or panic</u></b> .
3	Depression can change how you feel physically, your thinking, your emotions, and your behaviours: <b><u>Irritability and restlessness</u></b> <b><u>Aggressive</u></b> , abusive, or controlling behaviour.	Depression can change your body's feelings, ways of thinking, emotions, and behaviours: <b><u>Anxious</u></b> <b><u>Offensive</u></b> , abusive, or controlling behaviour.
4	The <b><u>Physical</u></b> Effects of Depression: <b><u>Psychomotor</u></b> <b><u>High blood pressure</u></b> <b><u>Medication side effects</u></b> <b><u>Coping</u></b> .	The <b><u>Physiological</u></b> Effects of Depression: <b><u>Mental movement</u></b> <b><u>Hypertension</u></b> <b><u>Drug side effects</u></b> <b><u>Response</u></b>
5	Clinical depression is marked by a <b><u>depressed mood</u></b> : Impaired concentration, indecisiveness Insomnia or <b><u>hypersomnia</u></b> (excessive sleeping) Restlessness or <b><u>feeling slowed down</u></b> Recurring thoughts of death or suicide	Clinical depression is characterised by <b><u>low mood</u></b> : <b><u>Inattention, indecision</u></b> Insomnia or <b><u>too much sleep</u></b> (too much sleep) Fidgeting or <b><u>feeling slow</u></b> Repeated thoughts of death or suicide

Zdroj: Wang et. al., 2021

Zdravotníci používající GT by mohli snížit potenciální riziko tak, že pacientům umožní číst překlady během ústních instrukcí; byli by pozorní na pravopis a gramatiku; a vyhýbali se složité gramatice, lékařskému žargonu a hovorové angličtině. (Khoong, Elaine C. et al 2019)

### 1.3.2. Strojový překlad v právu

Strojový překlad může mít negativní dopad také v právním kontextu. Například při získávání souhlasu k prohlídce v rámci policejních zásahů vedlo spoléhání se na strojový překlad k nedorozuměním a potenciálním porušením ústavních práv. V jednom případě, *United States v. Salas Antuna*, soud rozhodl, že i přesto, že fráze "¿Puedo buscar?" nebyla přesným překladem, policista na ni spoléhal v dobré víře, a tak byl důkaz získaný během

prohlídky přijat. Nicméně ve druhém případě, *United States v. Cruz-Zamora*, soud shledal, že dotyčný nemusel plně porozumět otázce přeložené přes Google Translate, a proto byly důkazy získané během prohlídky vyřazeny z procesu (*Texas Bar Journal* 2019: 94).

V dalším soudním sporu obvinila Společnost Novelty Textile, Inc., zabývající se textilním průmyslem, společnosti Xtaren a Windsor Fashions, Inc. z porušení copyrightu na design, který vlastní. V rámci sporu byly předloženy důkazy, mezi kterými byly i dokumenty přeložené pomocí Google Translate. Soud však zdůraznil, že překlady Google Translate nejsou dostatečně spolehlivé pro právní účely a odmítl je jako důkaz kvůli jejich nespolehlivosti. Tyto případy ukazují, jak mohou nedostatečné překlady ovlivnit výsledek soudního řízení, spravedlnost a důsledky pro obě strany (*Novelty Textile, INC. v. Windsor Fashions* 2013).

Dřívější výzkumy využití strojového překladu v právních kontextech často zkoumají přesnost překladu právních textů, tyto studie však mohou opomíjet praktické problémy hodnocení kvality překladů. Kritika se týká zejména toho, jak je hodnocení kvality překladu subjektivní a závislé na kontextu, což může vést k nepřesnostem v hodnocení systémů strojového překladu (Vieira, O'Hagan, O'Sullivan 2020:1524).

Madison Wahler navrhuje stanovení minimálních požadavků na přesnost jako možné řešení rizik spojených s používáním strojového překladu v právu. Přesto bývá v praxi obtížné dosáhnout dostatečné přesnosti, jelikož závisí na složitosti původního textu a na metodě hodnocení použité pro posouzení systému. Rostoucí závislost právníků na technologii strojového překladu s sebou nese riziko právních omylů způsobených chybami v překladech (Wahler 2018: 112, 113).

## **1.5 Etika a umělá inteligence**

Neurální strojový překlad může usnadnit komunikaci nad rámec starších metod, jeho použití má nicméně svoje důsledky. Vývoj technologií není eticky neutrální; odráží hodnoty těch, kdo je vyvíjejí. V následující části se budu zabývat etickými otázkami spojenými se strojovým překladem, počínaje sběrem dat a jeho opětovným využitím, až po to, jak strojový překladu koresponduje s hodnotami a kodexy překladatelů. Zvážím, zda mohou systémy odrážet "dobré" hodnoty a eliminovat zkreslení ve svých výstupech, a jaký přínos má strojový překlad pro budoucí vývoj společnosti.

Co je to etika? Etika se zabývá zkoumáním morálky, dobra a zla, a klade otázky o tom, jak nejlépe žít. Starověcí řečtí filozofové jako Sokrates a Aristotelés rozvíjeli pojmy dobrého života a ctností, které mají lidem umožnit dosáhnout osobního rozkvětu. Moderní etika nabízí různé přístupy k rozhodování o morálnosti akcí, jako je maximalizace štěstí pro většinu nebo jednání z čistých motivů. Aplikovaná etika se zaměřuje na konkrétní situace a často využívá etické kodexy, které řídí profesionální chování.

V kontextu strojového překladu je důležité zvážit etické aspekty vývoje a používání technologií, včetně možného vlivu na diverzitu a udržitelnost (Kenny 2022: 121, 122).

### **1.4.1. Vlastnictví dat a jejich využití**

Při zkoumání překladu jako výsledku kreativní činnosti překladatele je důležité si položit otázku, zda je možné tento výsledek chránit, a pokud ano, v jaké míře a za jakých podmínek. Obvykle bývají duševní výtvoři člověka chráněni zákonem, ale není tomu tak vždy bez výjimky. I když uznáme, že překladatelská práce je vždy tvůrčí činností, je důležité zvážit, zda každý překlad zasluhuje ochranu. „Existuje mezi původním a přeloženým dílem nějaký vztah, nebo jde o nezávislá díla? Je překlad vnímán jako kopie původního díla v jiném jazyce? A jaké postavení má překladatel jako tvůrce překladu?“ Těmito otázkami se zabývají odborníci nejen z oblasti filozofie, ale také práva. (Vargová, Varga 2023: 771)

Technologie poháněné daty, zejména neurální strojové překlady, vyžadují rozsáhlé množství dat pro trénink, což vyvolává otázky jejich etického využití. Například, i když je legální používat existující překladová data pro trénink systémů, může to být eticky sporné, pokud to vede k nevýhodám pro původní tvůrce dat, jako je snížení jejich odměn nebo neoprávněný prodej dat třetím stranám.

Další důležitou otázkou je vlastnictví dat a související práva. V mnoha případech mohou být data použita bez vědomí nebo explicitního souhlasu tvůrců, což vyvolává otázky o ochraně osobních údajů a autorských právech. Právní rámce často nejsou dostatečně vybaveny na to, aby řešily tyto moderní výzvy, což vyžaduje aktualizaci a rozšíření existujících zákonů, aby lépe reflektovaly potřeby a ochranu jak tvůrců, tak uživatelů překladových dat.

Dalšími potenciálními riziky může být zneužití dat, ztráta důvěry mezi překladateli a firmami, a další etická dilemata spojená s automatickým překladem, které mohou vést k

právním sporům a poškození reputace. Překladaelé se mohou setkat s komplikacemi, pokud jsou jejich data využita pro trénink strojových překladových systémů, což může být v rozporu s jejich etickými přesvědčeními. Představme si například následující situaci: Překladatel A mohl podepsat smlouvu umožňující využití jeho překladových dat k tréninku systémů strojového překladu bez vědomí konečného použití těchto dat. Pokud firma, pro kterou pracuje, rozhodne prodat tyto strojové překlady externím klientům, včetně výrobců zbraní, může se překladatel A nevědomky stát součástí dodavatelského řetězce pro zbrojní průmysl, i když s takovým použitím jeho práce nesouhlasí nebo o něm vůbec neví. Z tohoto důvodu je třeba zvážit zavedení minimálních standardů pro kvalifikaci překladatelů a pravidel pro používání technologie překladu, aby bylo zajištěno spravedlivé a etické využívání technologií.

V některých jurisdikcích je zaměstnavatel, který zaplatil za překlad, považován za právoplatného vlastníka dat, zatímco v jiných může dojít k převodu vlastnictví, což umožňuje jejich další využití. Často nedochází k vyžádání svolení pro opětovné použití získaných dat. Pokud překladaelé nemají možnost ovlivnit konečné využití své práce, může tato nejasnost v použití dat vést k etickým dilematům (Kenny 2022: 123, 124).

#### **1.4.2. Osobní údaje a soukromí**

Osobní údaje, které umožňují identifikaci jedince, jsou chráněny evropským nařízením GDPR (General Data Protection Regulation), které od roku 2018 stanovuje přísná pravidla pro sdílení a opětovné využití osobních údajů v rámci EU. GDPR ukládá organizacím povinnost hlásit porušení ochrany dat a stanovuje vysoké pokuty za jejich nedodržení, což může vést až k pokutám ve výši €20 milionů nebo 4 % z ročního obrátu. Tato regulace má za cíl zvýšit kybernetickou bezpečnost a omezit využívání serverů mimo území EU. Jakékoli sekundární využití osobních údajů by mělo být pokryto povolením uděleným pro původní užití, s některými výjimkami pro účely výzkumu. Vzhledem k tomu, že porušení ochrany dat může mít negativní dopad na důvěru spotřebitelů, tržní hodnotu a veřejné vnímání společnosti, některé společnosti se mohou snažit takové incidenty utajit. Transparentnost by však ideálně vedla k větší veřejné důvěře v organizaci. Dále je důležité si uvědomit, že ne všechna porušení ochrany dat jsou si rovna; některá mohou být dílem "etických hackerů" zaměstnaných organizací k identifikaci bezpečnostních rizik, zatímco jiná mohou mít za cíl získání dat pro osobní prospěch

„běžných“ hackerů. V kontextu rostoucího využívání webových platform pro překlad se překladatelé stávají stále více zranitelnými vůči zneužití svých osobních dat. Data z překladatelské činnosti tak mohou být propojována s dalšími osobními údaji a umožnit tak třetím stranám vyvozovat závěry o nich jako jednotlivcích nebo členech určité skupiny. Tyto platformy snižují kontrolu překladatelů nad jejich vlastními daty a umožňují dohled nad jejich pracovními aktivitami. V současné době probíhá snaha o anonymizaci překladových dat, tento proces je nicméně obtížný (Kenny 2022: 127).

Problém úniku dat se týká také takzvaných chatbotů, které jsou navrhovány tak, aby snižovaly potřebu lidské účasti v tradičních konverzích, například v e-shopech, bankovníctví, lékařských poradnách, psychoterapii a dalších oblastech, v souladu s pravidly GDPR. Zvláště během pandemie COVID-19 se ukázalo, že chatboti mohou být velmi užiteční a mohou šetřit čas, například na horkých linkách. V pandemických situacích však mohou lidé méně dbát na sdílení osobních údajů a odhalování svých strachů a obav. Chatboti pracují s daty uživatelů, a dokonce se z těchto dat 'učí'. To znamená, že by mohli být považováni za bezpečnostní problém, protože uživatelé většinou nevědí, jak jsou jejich citlivé osobní údaje zpracovávány, ukládány nebo dokonce sdíleny. Tato situace není považována za ideální a nové nařízení o ochraně osobních údajů, známé jako GDPR, také poukazuje na problémy spojené s bezpečností.

Obecné nařízení o ochraně osobních údajů (GDPR) je považováno za zlatý standard mezi předpisy o ochraně soukromí a slouží jako příklad pro ostatní státy. Chatboti mají přístup k široké škále osobních identifikačních informací a osobních údajů o svých uživateli. Pokud má chatbot přístup k osobním údajům uživatele, musí dodržovat předpisy GDPR. Článek 5 (e) GDPR stanovuje, že osobní údaje smí být uchovávány pouze po dobu nezbytně nutnou pro účely, pro které jsou zpracovávány.

Regulace GDPR zdůrazňuje, že každá technologie, která využívá osobní údaje uživatelů, musí udržovat přísné zabezpečení, aby nedošlo k porušení osobních práv a úniku osobních údajů. Každý chatbot musí mít zároveň „vhodná bezpečnostní opatření“ před neoprávněným vniknutím do systému.

Osobní údaje jakéhokoli typu tak musí plně zachovávat svou důvěrnost. Každý subjekt využívající chatboty a další technologie pracující s osobními a citlivými daty musí zajistit, aby k uloženým osobním údajům měl přístup pouze oprávněný uživatel.

Povinnosti správy dat pro procesy autorizace a autentizace mohou snížit bezpečnostní výzvy chatbotů. Mnoho společností, které vytvářejí platformy pro tvorbu chatbotů, přidalo příkazy jako „Request Download Data“, „Delete Personal Data“ nebo

„Change Personal Data“, které uživatelům poskytují určitou kontrolu nad jejich osobními údaji.

Další problematikou je provozování chatbotů na třetích stranách a komunikačních sítích, jako jsou Facebook Messenger, Viber, WhatsApp, Twitter, Facebook a LinkedIn. Tyto platformy mají rozdílné zásady týkající se toho, co patří uživateli a co společnosti, když uživatel využívá jejich služby. V roce 2018 bylo například na Facebooku aktivních více než 300 000 chatbotů, které mezi lidmi a podniky vyměnily 8 miliard zpráv. To je čtyřikrát více než v roce 2017. Tyto a podobné společnosti hodnotí data získaná z chatbotů, která mohou zahrnovat cookies, uživatelské preference, hlavní témata diskusí různých skupin, aktivitu, jazyk, pozitivní/negativní zprávy a další rysy, které lze z chatbotů extrahovat a prodat jiným zákazníkům. Všechny tyto prvky spadají do kategorie osobních identifikačních informací a jsou velmi cenné pro personalizaci a vyhledávání informací.

Nedávno Facebook WhatsApp změnil své podmínky a zásady ochrany osobních údajů tak, aby umožnil propojení WhatsApp s dalšími společnostmi Facebooku a využití některých informací shromážděných z WhatsApp. To znamená, že obchodní zprávy mohou být využívány pro komerční účely a některé z nich jsou také ukládány na serverech Facebooku. Jak bylo zmíněno na začátku tohoto odstavce, mnoho obchodních zpráv je vytvářeno obchodními chatboty. Tyto aspekty by měly být zváženy při posuzování ochrany soukromí a bezpečnosti dat u chatbotů provozovaných na platformách třetích stran (Hasal et al. 2021:7, 8, 9).

### **1.4.3. Předsudky vůči pohlaví ve strojovém překladu**

V současné době narůstá znepokojení mezi akademiky, výzkumnými laboratořemi a hlavními komerčními médii ohledně jevu zvaného „strojová zaujatost“. Tento jev se týká statistických modelů, které nevědomky reflektují společenské nerovnosti, jako jsou pohlavní a rasové předsudky. V poslední době bylo nahlášeno několik případů, kdy nástroje umělé inteligence projevily škodlivou zaujatost vůči určitým menšinám, včetně případů rasově zaujatých předpokladů kriminálního chování, iPhone X od Apple selhal ve schopnosti rozlišit mezi dvěma Asiaty a skandálu s Google Photos, kde byli lidé s černou pletí mylně klasifikováni jako gorily.

Ve studii z roku 2019 byl sestaven obsáhlý seznam pracovních pozic z U.S. Bureau of Labor Statistics (BLS), který byl použit k vytvoření vět typu "He/She is an Engineer"

(kde "Engineer" je nahrazen relevantní pracovní pozicí) ve 12 různých rodově neutrálních jazycích, jako jsou maďarština, čínština nebo jorubština. Tyto věty byly přeloženy do angličtiny pomocí Google Translate. Tyto data o četnosti výskytu mužských, ženských a rodově neutrálních zájmen v překladech byly následně analyzovány. Analýza ukázala, že Google Translate má silnou tendenci používat primárně mužská zájmena, zejména v oborech s nerovnoměrným zastoupením pohlaví nebo stereotypy, jako jsou pozice ve vědeckých, technologických, inženýrských a matematických oblastech.

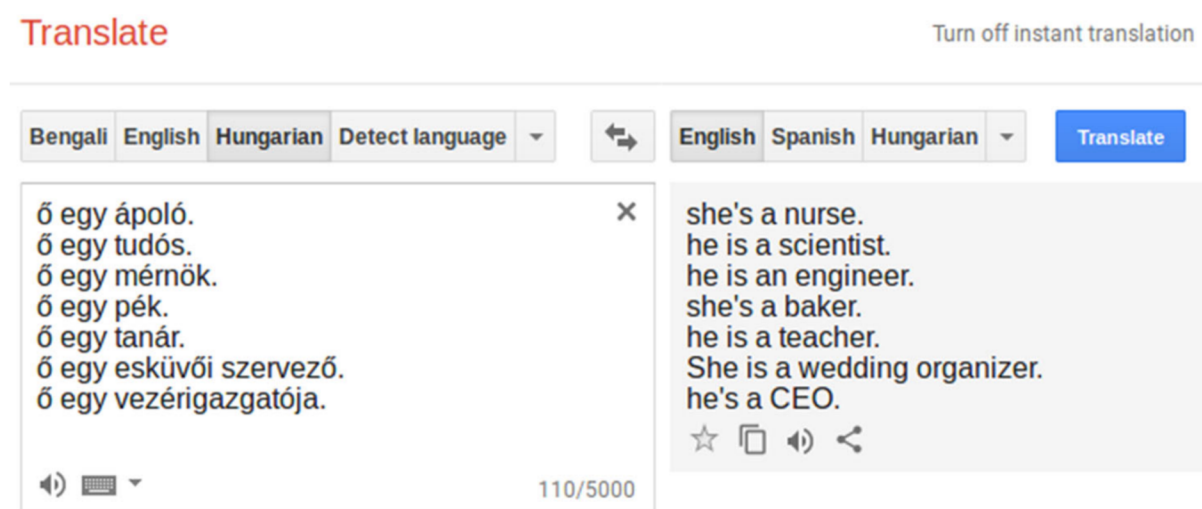
Na základě těchto dat bylo zjištěno, že Google Translate neodráží reálné rozložení žen v pracovním prostředí. Z výsledků experimentu vyplývá, že Google Translate upřednostňuje mužská zájmena mnohem častěji, než by odpovídalo demografickým údajům.

Google Translate patří mezi největší veřejně dostupné nástroje pro strojový překlad, s denním počtem uživatelů přesahujícím 200 milionů. Tento nástroj původně využíval data z jednání OSN a Evropského parlamentu a od roku 2014 získává obsah také od svých uživatelů prostřednictvím iniciativy „Translate Community“. V poslední době se objevují obavy ohledně genderových asymetrií v jeho překladovém mechanismu, což někteří kritizují jako "sexistické". Tyto obavy mají oporu ve vědeckých studiích, které ukázaly, že techniky pro vkládání slov mají tendenci reprodukovat genderové stereotypy. Pro odstranění těchto předsudků byl navržen mechanismus, který úspěšně snížil počet stereotypních analogií z 19 % na 6 %, aniž by to významně ovlivnilo funkčnost těchto technik. Tato pozitivní zjištění naznačují možnost aplikace podobných technik pro eliminaci genderového zkreslení ve výstupech z Google Translate.

Statistický překladový nástroj by měl co nejvíce odrážet stávající sociální nerovnosti, protože překládá z příkladů vytvořených společnostmi a nevyhnutelně tak přejímá určité existující zkreslení. Diskutuje se o tom, že jazyk může ovlivňovat vnímání světa člověka, a proto může gramatické rozlišování pohlaví v některých jazycích posilovat genderové předsudky. V jazycích jako je angličtina, kde je možné dosáhnout genderové neutrality, by překladové nástroje měly tuto neutralitu udržet a nevytvářet automaticky mužské nebo ženské varianty (Prates et. al., 2019).



Obr. č. 1 (Překlad pracovních pozic z maďarštiny)



Zdroj: Prates et. al., 2019

## 1.6 Hodnocení a post-editace strojového překladu

Technologický pokrok a konkurenční tlak vedou k potřebě přeložit mnohem více obsahu než kdykoli předtím, a to nejen rychleji, ale také s výrazně nižšími náklady. Je proto nezbytné hledat způsoby, jak zajistit, aby byl proces překladu co nejrychlejší, nejpresnější a nejefektivnější – ať už za použití softwarových nástrojů, nebo bez nich. V tomto kontextu získává klíčový význam role strojového překladu a následné post-editace.

Podobně významnou změnou prošel i výzkum překladu: badatelé se odklánějí od zkoumání koncepčních a pedagogických otázek a více se soustředí na sběr systematických empirických dat z reálných překladatelských úkolů, které zahrnují průmyslové a kognitivní překladatelské procesy. Díky tomu dnes působí více výzkumníků, jsou k dispozici pokročilejší výzkumné nástroje a také množství podrobnějších dat ve srovnání se stavem před pouhými deseti lety.

Podle Whitea a O'Connella bylo hodnocení strojového překladu vždy mimořádně náročné. Toto tvrzení platí dodnes. Hodnocení má v oblasti strojového překladu zásadní význam, protože pomáhá zjistit úroveň kvality výstupu systému, jak si vede ve srovnání s ostatními systémy a zda je dostatečně kvalitní pro základní pochopení nebo další úpravy. Hodnocení je však také složité, a to kvůli subjektivitě, času a nákladům, které s tím souvisí. Z kognitivního hlediska je hodnocení strojového překladu silně závislé na lidském vnímání, což je zároveň výhodou i slabinou. Bylo zjištěno, že hodnocení prováděné lidmi

postrádá shodu mezi jednotlivými hodnotiteli, ale i přes to zůstává zásadní, protože hlavními uživateli výstupů jsou lidé a dokážou přesně posoudit závažnost chyb (O'Brien 2017: 314).

S rostoucím objemem textu, který je třeba překládat rychleji a za nižší cenu, se post-editace strojového překladu stala nezbytnou praxí. I přes to, že je to mentálně náročná úloha, často se o ní uvažuje jen jako o opravě chyb vytvořených strojovým překladem, místo aby byla chápána jako tradiční překlad. Tento pohled vede k domněnce, že post-editace je jednodušší než samotný překlad, což se odráží v nižších cenových sazbách.

Rostoucí potřeba po překladech vedla ke vzniku dvou úrovní post-editace, rozlišovaných jako lehká a úplná. Lehčí varianta post-editace zahrnuje provedení pouze nezbytných úprav co nejrychleji, zatímco úplná post-editace opravuje všechny chyby a vyžaduje více času. Obě varianty by však měly být rychlejší než tradiční překlad bez podpory překladových nástrojů.

Standard ISO 18857 definuje lehkou post-editaci jako proces, při kterém se provádí pouze nezbytné úpravy, které zajišťují srozumitelnost textu bez úsilí vytvořit překlad srovnatelný s lidským překladem. Úplná post-editace se pak snaží dosáhnout kvality plně srovnatelné s tradičním překladem.

### Obr. č. 3 (Rozdíly mezi lehkou a úplnou post-editací)

Lehká post-editace	Úplná post-editace
Cílem je sémanticky správný překlad.	Cílem je gramaticky, syntakticky a sémanticky správný překlad.
Zajistit, aby klíčová terminologie byla přeložena správně a nepřeložené termíny patřily na klientův seznam "Nepřekládat".	Zajistit, aby klíčová terminologie byla správně přeložena a nepřeložené termíny patřily na klientův seznam "Nepřekládat".
Zajistit, aby nedošlo k neúmyslnému přidání nebo vynechání informací.	Zajistit, aby nedošlo k neúmyslnému přidání nebo vynechání informací.
Opravit jakýkoli urážlivý, nevhodný nebo kulturně nepřijatelný obsah.	Opravit jakýkoli urážlivý, nevhodný nebo kulturně nepřijatelný obsah.
Použít co nejvíce původního výstupu MT.	Použít co nejvíce původního výstupu MT.
Základní pravidla pravopisu platí.	Uplatnit základní pravidla pravopisu, interpunkce a dělení slov.
Není třeba provádět stylistické úpravy.	
Není třeba restrukturalizovat věty pouze za účelem zlepšení přirozeného toku textu.	
	Zajistit správné formátování.

Zdroj: (Kenny 2022: 109)

Hlavním cílem post-editace je zajistit srozumitelnost výsledného textu, přesnou shodu mezi zdrojovým a cílovým textem a dodržení požadavků stanovených poskytovatelem překladatelských služeb. Aby se tyto cíle naplnily, je nutné dodržovat kritéria, jako jsou konzistentní terminologie a přesnost terminologických pojmů, dodržování jazykových standardů, správné formátování, zaměření na cílové publikum a účel překladu.

Rozdíl mezi lehkou a úplnou post-editací může být subjektivní a liší se v závislosti na organizaci nebo projektu.

Co největší využití z výstupu strojového překladu je klíčové, i když mohou mít překladatelé nutkání ho ignorovat a překládat přímo ze zdrojového textu. Neuronový strojový překlad však zlepšil kvalitu surového výstupu natolik, že je často užitečný.

Úrovně post-editace jsou také spojeny s různými úrovněmi kvality. Lehkou post-editaci lze vnímat jako snahu o "dostatečně dobrou kvalitu", zatímco úplná post-editace je srovnatelná s lidským překladem. (Kenny 2022: 107,108)

Post-editace může být prováděna v libovolném textovém editoru, nejčastěji se však používají nástroje s překladovou pamětí v CAT (computer-aided translation) nástrojích. Ty kombinují dříve přeložené části s návrhy ze strojového překladu. Překladatel může upravit nastavení, aby viděl návrhy z obou systémů současně a vybral ten nejlepší. Přesná shoda z překladové paměti je obvykle spolehlivější než návrhy od strojového překladu, nicméně pokročilé systémy strojového překladu mohou být užitečnější než částečné shody z překladové paměti. Překladatel tak získává více možností a podpory při práci.

Tento přístup však může zvýšit kognitivní zátěž, jelikož překladatel zpracovává velké množství informací (návrhy z překladové paměti i strojového překladu). Kombinace post-editace, revize a překladu tak může být náročná, zejména pod časovým tlakem.

Kvalitní post-editor by tak měl být především dobrým překladatelem s rozvinutými dovednostmi, který by měl umět rozpoznat a opravit chyby ve výstupu strojového překladu, pracovat rychle a efektivně podle pokynů, aniž by prováděl nadbytečné úpravy.

Měl by také mít pozitivní postoj ke strojovému překladu. Pokud překladatel tuto technologii odmítá, může být v pokušení ignorovat nebo mazat jeho návrhu, což vede k větší časové náročnosti a vyšším nákladům.

V poslední dekádě se rozvíjí výuka zaměřená na strojový překlad a post-editaci. Univerzitní překladatelské programy integrují tyto technologie do svých kurzů, aby studenti pochopili silné a slabé stránky strojového překladu, naučili se ho hodnotit a

efektivně post-editovat. Zásadní je také naučit překladatele, kdy a jak využít MT optimálně (Kenny 2022: 118).

## 2 Praktická část

### 2.1 Charakteristika překládaného textu

Odborný text o umělé inteligenci je podrobnou publikací, která poskytuje komplexní přehled o různých aspektech umělé inteligence (UI). Tento text zahrnuje množství témat, počínaje základními definicemi a historií UI, přes aktuální využití v praxi, až po diskusi o etických a právních dilematech. Publikace systematicky analyzuje technologický pokrok a jeho dopad na mezinárodní vztahy a sociální struktury, zkoumá aplikaci UI ve vojenském sektoru a jeho potenciální nebezpečí. Podrobně se věnuje rozlišení mezi "slabou" a "silnou" UI a popisuje, jak mohou tyto technologie proměnit pracovní trh, zdravotnictví, domácí život a mnoho dalších oblastí. Text také zkoumá koncepty jako je strojové učení a autonomní systémy, a přináší statistická data o vývoji a implementaci UI v různých průmyslových odvětvích. Z důvodu rozsáhlosti se moje analýza omezí na vybrané kapitoly, konkrétně na ty, které se týkají vývoje, současné aplikace a mezinárodních aspektů umělé inteligence, a také na kapitoly zabývající se etickými otázkami a budoucími perspektivami UI. V rámci této práce budu provádět teoreticko-praktickou analýzu vybraných částí textu, abych poskytl detailní pohled na současný stav a možný budoucí vývoj UI, její význam pro budoucnost společnosti a přínos pro technologický pokrok.

### 2.2 Analýza překládaného textu

V následující části se budu věnovat formální a obsahové analýze překládaného textu s názvem *„Международные и социальные последствия использования технологий искусственного интеллекта.“* V první části se zaměřím na obecnou charakteristiku a kompozici textu, z jakých částí se skládá, o čem jednotlivé části pojednávají a jaký je hlavní cíl originálního textu. V další fázi přejdu k detailní syntaktické, morfologické a lexikální analýze, kde se zaměřím na lingvistické aspekty překládaného textu, jako je složení vět, styl psaní, členění textu, a terminologie.

#### 2.2.1. Kompozice textu

Překládaný text komplexně rozpracovává důsledky implementace umělé inteligence (UI) v mezinárodním a sociálním kontextu. V horizontální rovině je dokument

organizován do členěných kapitol, které jsou dále rozděleny do logické do podkapitol a odstavců, umožňujících ucelený pohled na danou problematiku.

V úvodu je nastíněn obecný popis fenoménu umělé inteligence a některých probíraných témat pro její lepší pochopení. Dále text pokračuje popisem historie a vývoje, rozebírá možnosti současného i budoucí využití, a zabývá se možnými riziky a eticko-právními otázkami. Kapitoly jsou strukturované tak, aby navazovaly v přirozeném postupu, od obecných přehledů k detailnějším rozborům specifických témat. Vertikálně je text doplněn odkazy na zdroje formou poznámek pod čarou. V textu se nachází také několik ilustrací, tabulek a grafů, které pomáhají lepšímu celkovému porozumění a hlubšímu pochopení prezentovaných dat.

Dokument dále zahrnuje přehledné shrnutí perspektiv UI v oblasti mezinárodních vztahů a jejího vlivu na společenský život, čímž poskytuje čtenáři možnost získat komplexní přehled o současném a budoucím směřování této rychle se vyvíjející technologické oblasti.

## 2.2.2. Syntaktická analýza

Překládaný text je z odborné technické oblasti, a proto je z velké části tvořen odbornými termíny, složitými větami, a dlouhými souvětími, které byly často obtížně pochopitelné. Text je psán spisovným odborným jazykem.

Původce děje je vyjádřen všeobecně, neurčitě nebo vůbec. Zde je několik příkladů:

*Под понятием «искусственный интеллект» (ИИ), как правило, понимают компьютерную программу...*

*К концу 2010-х годов можно выделить следующие магистральные направления исследований...*

*На текущий момент под исследованиями в области искусственного интеллекта сейчас принято...*

V textu můžeme často vidět použití pasivních konstrukcí, sloužících k objektivnímu podání informací.

Například věta „*В публикациях также рассматриваются этико-юридические аспекты применения технологий ИИ...*“ klade důraz na samotný proces zkoumání, aniž by byl zmíněn subjekt, který dané zkoumání provádí.

Дале text obsahuje složité větné konstrukce spolu s větami vedlejšími, které slouží k rozšíření a detailnějšímu rozpracování hlavní myšlenky. Například, „Искусственный интеллект, который...“ vede k rozvíjení definice umělé inteligence s podrobnými specifikacemi jejích charakteristik. Dalším příkladem složité větné konstrukce, složené z několika vět vedlejších, může být následující odstavec:

*„В-пятых, угрозу для общества представляет переход к новому уровню развития производственных отношений в капиталистическом (либо тоталитарном) обществе, когда более малочисленная часть населения получает возможность контролировать материальное производство, исключая из него подавляющую часть населения за счет еще большей автоматизации, что может приводить к еще большему социальному расслоению, снижению эффективности «социальных лифтов» и увеличению массы «лишних людей» с соответствующими социальными последствиями.*

Dalším znakem, který se v textu často vyskytuje, jsou infinitivní konstrukce, které slouží k ekonomičtějším a formálnějším vyjádření dané informace:

*„...позволяющих вычислительным системам действовать...“*

*„...в мощные вычислительные системы способные не только обрабатывать графические данные...“*

V textu se také vyskytují polovětné konstrukce, jako přechodníky (деепричастия), a příčestí (причастия) které slouží především ke kondenzaci a stručnému vyjádření informací.

Přechodníky:

*„...разработку технологий, позволяющих вычислительным системам...“*

*„...решение проблем, требующих когнитивного напряжения...“*

*„...объединившая в себе огромный массив знаний...“*

Пříčestí:

*„...сократив время обработки накопленных массивов...“*

*„...Построив и натренировав нейросеть с подобной архитектурой...“*

Pro ruštinu je na rozdíl od češtiny typický inverzní slovosled, který slouží k zdůraznění a vyzdvižení určité informace.

*„...Так, известный большинству пользователей переводчик компании*

*Google работает... “*

V tomto příkladu je zdůrazněn výraz „переводчик“. „...Даже для кажущихся всемогущими компьютерных алгоритмов... “

### 2.2.3. Морфологická analýza

Z morfologického pohledu se textu se nejčastěji vyskytují substantiva, což je charakteristické odborný styl a slouží k přesné specifikaci konceptů a technologií. V textu je také spousta akronymů, jako například „ИИ“ (UI – Umělá inteligence), „НИОКР“ (*Научно-исследовательские и опытно-конструкторские работы* – Věda a výzkum), a termínů jako „искусственный интеллект“ (umělá inteligence), nebo „машинное обучение“ (strojové učení). Substantiva často tvořena ze složených termínů, jako například „нейронные сети“, „автономные системы“ (autonomní systémy), nebo „автономный агент“ (autonomní agent), což ukazuje na složitost tématu umělé inteligence.

Dále se objevují slova s koncovkami typickými pro odborný styl, jako „-ция“ ve slovech „классификация“ (klasifikace) nebo „глобализация“ (globalizace), které naznačují procesy nebo akce spojené s daným pojmem. Často se také vyskytuje koncovka „-ение“, jako například ve slovech „использование“, a „обучение“. Mezi dalšími častými koncovkami se vyskytuje koncovka „-ство“, například ve slovech „сообщество“, „производство“, „лидерство“ a další.

Adjektiva jsou v textu používána k popisu různých technologií a k upřesnění jejich vlastností, například „автономный“ (autonomní), „интеллектуальный“ (inteligentní), což napomáhá vytvoření detailního obrazu o technologiích UI. V textu se objevují i složené adjektivní formy, které popisují specifické charakteristiky systémů umělé inteligence, jako je „сверхчеловеческий“ (nadlidský), „высокопроизводительные“ (vysokovýkonné). Mezi nejčastější koncovku vyskytující se u adjektiv patří koncovka „-ный“, například „международный“, „потенциальный“, „эффективный“, „искусственный“.

V textu se také často setkáváme s množstvím krátkých přídavných jmen, která popisují stálé vlastnosti objektů nebo jevů. Mezi ně patří slova jako „конкуренция уникальна“ a „угрозами разного масштаба, которые актуальны“.

Slovesa v textu často ukazují na akce, procesy nebo schopnosti spojené s umělou inteligencí, jako je „обучать“ (učit), „распознавать“ (rozpoznávat), a „принимать решения“ (přijímat rozhodnutí), což vystihuje dynamický charakter této technologie.



Zajímavým jevem v textu je využití složených výrazů a odborných frází, které ukazují na multidisciplinární povahu oblasti umělé inteligence, jako jsou „*биологически-вдохновленные когнитивные архитектуры*“ (biologicky inspirované kognitivní architektury) a „*стелс-технологии*“ (stealth technologie). Tyto fráze poukazují v rámci vývoje umělé inteligence na propojení informatiky, kognitivní vědy a neurologie.

#### 2.2.4. Lexikální analýza

Po lexikální stránce text obsahuje širokou škálu terminologie spojené s umělou inteligencí a dalšími technologiemi, včetně odborných termínů jako „*слабый ИИ*“ (slabá umělá inteligence), „*младенческий тест Тьюринга*“ (dětský Turingův test) a „*нанотехнология*“ (nanotechnologie).

Mezi slova, která se v textu vyskytují nejfrekventovaněji, patří technické termíny jako „*нейросети*“ (neuronové sítě), „*алгоритмы*“ (algoritmy), a „*искусственный интеллект*“ (umělá inteligence). Dále jsou zastoupena slova abstraktního charakteru s nominativním charakterem, například „*развитие*“ (rozvoj), „*применение*“ (použití), „*влияние*“ (vliv).

V textu nalezneme termíny specifické pro oblast umělé inteligence, jako jsou „*сверхчеловеческая уровень*“ (nadlidská úroveň), „*общий ИИ*“ (obecná umělá inteligence) a „*роботизированные*“ (robotizované). Složené termíny, jako „*автономное планирование*“ (autonomní plánování) a „*глубокое обучение*“ (hluboké učení), ukazují na specifické aplikace AI technologií.

V oblasti spojení a ustálených výrazů se setkáváme s frázemi jako „*необходимо отметить.*“ (..je důležité poznamenat..) nebo „*ожидается*“ (očekává se). Text také reflektuje mezinárodní a sociální důsledky používání AI technologií, což je patrné z opakovaného užití spojení „*международные отношения*“ (mezinárodní vztahy) a „*общественная жизнь*“ (veřejný život). Z tohoto pohledu se také často opakuje slovní spojení „*социальные последствия*“ (sociální důsledky).

V kontextu právní a etické problematiky se objevují termíny jako „*этико-юридические аспекты*“ (eticko-právní aspekty), což poukazuje na výzvy v oblasti regulací a morálky spojené s integrací UI do společnosti.

Lexikální analýza poukazuje, že slovní zásoba v textu je charakteristická pro odborný styl, který se vyznačuje použitím specifických technických termínů a složených

výrazů, které zdůrazňují různé aspekty umělé inteligence a jejího vlivu na mezinárodní vztahy a společnost.

### **2.2.5.      Obsahová analýza**

Po obsahová stránce je text rozdělen do 8 hlavních kapitol, zaměřených na různé otázky spojené s problematikou umělé inteligence jak v oblasti sociální, co se dopadů na společnost týká, tak v oblasti vojenské z hlediska potenciálního nebezpečí pro mezinárodní bezpečnost a stabilitu. První dvě kapitoly „*Искусственный интеллект: время слабых*“ (Umělá inteligence – Doba slabých) a „*Искусственный интеллект – благо или угроза для человечества?*“ (Umělá inteligence – prospěch nebo hrozba pro společnost) jsou primárně zaměřené na historii UI, její vývoj a současný stav. Součástí těchto kapitol je také základní rozdělení existujících i zatím neexistujících druhů umělé inteligence s popisem a příklady jejich potenciálu jak konstruktivního, tak destruktivního. Nejdelsí kapitolou v textu je „*ИИ-национализм и ИИ-национализация*“ (Nacionalismus a nacionalizace UI), která z důvodu rozsahu nebyla součástí překladu. Tato kapitola popisuje nové termíny spojené s vývojem umělé inteligence. Jde o UI – Nacionalismus, který v oblasti strategií spojených s UI deklaruje ekonomickou a vojenskou prioritizaci zájmů své země a UI – Nacionalizace neboli proces, ve kterém státy rozvíjejí a implementují své vlastní strategie umělé inteligence s cílem zvýšit svou domácí bezpečnost, hospodářskou sílu a technologickou suverenitu za spolupráce státu a soukromých společností. Ve zbývajících dvou kapitolách zaměřených na téma umělé inteligence s názvy „*Дивный новый мир без работы*“ a „*Этические и правовые вопросы искусственного интеллекта*“ se práce věnuje potenciální budoucnosti vývoje umělé inteligence v pracovní sféře, kde popisuje revoluční změnu určitých pracovních pozic, zaměřuje se na profese, které mají vyšší pravděpodobnost nahrazení a automatizace a věnuje se také profesím, které naopak čeká rozkvět. Současně se text zabývá etickými a právními otázkami spojenými s rostoucí autonomií umělé inteligence, jako jsou otázky odpovědnosti za chyby v rozhodování UI a jak tyto problémy mohou být právně adresovány. V neposlední řadě práce shrnuje právní iniciativy jednotlivých států a jejich snahy směřující k vytvoření vhodných podmínek pro rozvoj a rozšiřování technologií umělé inteligence.

## 3 Praktická část – překlad

### Úvod

Umělá inteligence (UI), která ještě nedávno existovala pouze v science fiction, je dnes jednou z nejperspektivnějších a rychle se rozvíjejících technologií. Technologie omezené nebo „slabé“ umělé inteligence jsou již aktivně používány v nejrůznějších oblastech, od mobilních telefonů a domácí elektroniky až po vojenskou výrobu.

Zároveň existuje málo vědních oblastí, které jsou tak zahaleny mýty, jako je UI. Díky popkultuře je umělá inteligence ve sdělovacích prostředcích často prezentována téměř jako Pandořina skříňka, jejíž otevření by nevyhnutelně uvedlo lidstvo na pokraj apokalypsy. Ačkoli je tento alarmismus neopodstatněný, již dnes probíhá vývoj „silné“ umělé inteligence, schopné samostatně činit vědomá rozhodnutí.

Perspektiva vytvoření takové technologie významně zpochybňuje nejen současný systém globální dělby práce, ale také stávající světový řád a architekturu mezinárodní bezpečnosti. V kontextu zhoršujících se rozporů a krize důvěry mezi velmocemi existuje reálné nebezpečí vzniku nového závodu ve zbrojení v oblasti technologií umělé inteligence. Zatímco kontrola nad jadernými, chemickými a biologickými zbraněmi hromadného ničení je regulována příslušnými mezinárodními dohodami a úmluvami, vývoj vojenských technologií umělé inteligence zůstává v „šedé zóně“ mezinárodního práva.

Zároveň je dnes zřejmé, že umělá inteligence bude čím dál tím více využívána ve vojenské a civilní sféře, včetně otázek zajištění strategické stability. Pokušení získat nové dokonalejší zbraně a dostat se vpřed v technologickém závodu je příliš velké, než aby země mohly seriózně zohlednit společenské důsledky.

Jedním z problémů, kterým čelí výzkumníci, je velká pojmová a terminologická zmatenost v oblasti UI. Tento dokument systematizuje stávající množství znalostí o technologiích umělé inteligence a hodnotí perspektivy jejího vývoje.

V centru analýzy je možný dopad UI a s ní souvisejících technologií (strojové učení, autonomní zařízení) na mezinárodní vztahy a různé sféry společenského života. Publikace také zkoumá eticko-právní aspekty používání technologií UI.

Dokument Ruské rady pro mezinárodní vztahy (RRMV) obsahuje analytické materiály, které připravili specialisté v oblastech umělé inteligence, strojového učení, autonomních systémů, stejně jako právníci a sociologové. Materiál má přispět k veřejné diskusi o otázkách umělé inteligence a možných důsledcích použití této technologie.

## **Umělá inteligence: Doba slabých**

Šeftelovič D. R.

Pod pojmem „umělá inteligence“ (UI) se obvykle rozumí počítačový program, tedy algoritmus, schopný řešit úkoly, které dokáže řešit mozek dospělého člověka. V Mezinárodním terminologickém slovníku umělé inteligence je pojem „umělá inteligence“ definován jako oblast znalostí, zkoumající vývoj technologií, které umožňují výpočetním systémům jednat způsobem, který připomíná racionální chování, včetně chování člověka. Je třeba poznamenat, že se jedná o fenomenologickou definici, která ponechává podrobnou charakteristiku pojmů „inteligence“, „racionality“ na filozofii; v podmínkách nedostatku znalostí o mozku a kognitivním aparátu biologických systémů se pojem UI nejde přesněji, matematicky formalizovat.

V současné době se pod výzkumem v oblasti umělé inteligence obvykle rozumí algoritmické řešení problémů, vyžadujících kognitivní úsilí. Takovými problémy byly (a částečně stále jsou) hry: šachy nebo Go, rozpoznávání rukopisu, strojový překlad, tvůrčí činnost. Ve veřejném vědomí se každý z těchto problémů původně jevil jako poslední překážka k vytvoření „pravé“ UI, schopné nahradit člověka ve všech oblastech jeho působení. Ve skutečnosti se ukázalo, že, za prvé, naučit počítač hrát šachy je mnohem jednodušší než ho naučit hrát fotbal, a za druhé, i když víte, jak naučit počítač hrát fotbal, je těžké přenést tuto zkušenost na problém strojového překladu. Proto po počátečním nadšení vědecká komunita rozdělila pojem UI na kategorie silné a slabé UI. Pod slabou umělou inteligencí se rozumí algoritmus, schopný řešit úzce specializované problémy (šachy), zatímco silná umělá inteligence má širší spektrum úkolů; ideální silná umělá

inteligence je schopna alespoň toho, co dospělý člověk (například dělat logické závěry a plánovat činnosti). Místo pojmu silná UI se v literatuře také používá umělá obecná inteligence (Artificial General Intelligence, AGI).

Je třeba poznamenat, že klasifikace algoritmických problémů na problémy silné a slabé umělé inteligence není historickou konstantou; tak například na začátku roku 2016 se mělo za to, že pro hraní hry Go je potřeba silná umělá inteligence. Po impozantním vítězství algoritmu AlphaGo nad tehdejším světovým šampionem v březnu 2016 se hra Go stala problémem slabé umělé inteligence.

Ke konci 10. let 21. století lze v oblasti výzkumu algoritmů umělé inteligence vyčlenit následující hlavní směry:

#### **Strojové učení a rozpoznávání vzorů:**

Hledání vzorců v datech, například klasifikace objektů na fotografii do kategorií jako "pozadí", "člověk", "auto", "budova", "rostlina".

#### **Plánování a logické usuzování:**

Důkaz logických tvrzení a plánování akcí pro dosažení určitého cíle s využitím logických zákonitostí, které umožňují tohoto cíle dosáhnout. Příkladem je syntéza dat ze senzorů pro hodnocení dopravní situace a adekvátní řízení vozidla.

#### **Expertní systémy:**

Systematizace dat spolu s logickými souvislostmi, zobrazování znalostí, odpovědi na sémantické dotazy typu "Jaký je podíl cen za energie ve výrobních nákladech letadla MS-21?".

#### **Současný stav a perspektivy:**

Problém vytvoření silné umělé inteligence zatím nebyl vyřešen, a vědecká komunita se po takzvané "zimě umělé inteligence", která nastala v polovině 80. let 20. století kvůli přehnaným očekáváním, k tématu umělé inteligence výrazně ochladila. Z jedné strany to vedlo ke zklamání potenciálních uživatelů a ze strany druhé k nedostatečné výkonnosti softwarových systémů. Na začátku 21. století se však podařilo dosáhnout značných úspěchů ve specializovanějších úlohách. To je především dáno neustálým vývojem

výpočetní techniky: zatímco v roce 2001 byl "vlajkovým" procesorem Intel Pentium III, který zvládal 1,4 miliardy aritmetických operací za sekundu, deset let poté, v roce 2011, byl aktuálním procesorem Intel Core i5, který vykonával již 120 miliard operací za sekundu, tedy téměř 100krát více.

V mnohém bylo motorem růstu odvětví počítačové grafiky a videoher: ve snaze o stále realističtější grafiku se grafické karty z periferního zařízení proměnily v silné výpočetní systémy, schopné nejen zpracovávat grafická data, ale také provádět libovolné (byť ne příliš složité) paralelní výpočty. V roce 2011 tak výkonnost grafické karty Radeon HD 6970 dosahovala hodnoty 675 miliard operací za sekundu.

Druhým důležitým faktorem rozvoje odvětví se stala digitalizace a ruční klasifikace textů, fotografií a zvukových záznamů, stejně jako budování univerzálních databází znalostí v digitálním formátu. Dostupnost velkých objemů kvalitně zpracovaných dat umožnila trénovat algoritmy strojového učení na velkých vzorcích. Tím došlo k nárůstu přesnosti klasifikace, a strojový překlad se proměnil z hrubého nástroje na systém s širokým využitím. Například známý překladač společnosti Google funguje díky obrovské databázi paralelních textů, tedy textů identického významu v různých jazycích, které jsou doplňovány uživateli, což umožňuje systém trénovat a zlepšovat kvalitu překladu přímo během práce.

Díky zkrácení doby zpracování nahromaděných datových souborů umožnily oba tyto faktory řešit v krátkém čase stále složitější úkoly. Na začátku roku 2018 umělá inteligence dosáhla několika významných úspěchů zviditelněných médií, a pronikla tak do různých odvětví ekonomiky. Nicméně považovat umělou inteligenci za zázračné řešení, prostředek, schopný vyřešit všechny problémy lidstva, by bylo chybou.

## **Úspěchy UI**

### **"Expertní systémy a rozpoznávání textů: vlastní hra IBM Watson.**

Patrně největší pozornost veřejnosti přitáhl expertní systém IBM Watson, který kombinoval obrovskou databázi znalostí, tedy data se sémantickými vazbami mezi nimi, a schopnost zpracovávat dotazy v anglickém jazyce k této databázi znalostí. V roce 2011 Watson dosáhl pozoruhodného úspěchu, když porazil stávající šampióny v televizní kvízové hře „Jeopardy!“. Tento úspěch umožnil společnosti IBM úspěšně proniknout na

trh expertních systémů, kde názorně demonstroval schopnosti systému Watson zpracovávat a strukturovat informace."

Neuronové sítě: Vítězství v Go s Google DeepMind

V mnohém těžila z růstu výpočetní síly technologie umělých neuronových sítí, známá již od poloviny 40. let 20. století. Současné technologie umožňují trénovat velké neuronové sítě, přičemž velikost sítě přímo souvisí se složitostí řešených úloh. Charakteristickým rysem těchto systémů hlubokého učení je struktura sítě, která umožňuje nejprve rozpoznávat lokální detaily vstupních dat (například rozdíly v barvě mezi sousedními pixely) a pak také zpracování jejich obecných vlastností (například linie a tvary). Po vytvoření a zaučení takové neuronové sítě se inženýrům z Googlu podařilo vytvořit počítačový algoritmus pro hru Go, který neočekávaně porazil mistra světa. Podobných výsledků dosáhly i jiné výzkumné skupiny, které naučily neuronovou síť hrát počítačové hry jako je StarCraft II.

### **Praktické použití UI**

K začátku roku 2018 vyšlo strojové učení, rozpoznávání obrazů a autonomní plánování z výzkumných laboratoří do světa komerčních aplikací. Prvními uživateli nové technologie se staly armády na obou stranách železné opony, které se začaly zajímat o řešení plánovacích úloh od 50. let. Téměř současně s nimi začali plánovací úlohy řešit i ekonomové. Uvedeme několik příkladů použití algoritmů umělé inteligence.

- IBM nabízí služby již zmíněného systému Watson různým odvětvím: lékařům pro diagnostiku symptomů pacienta a doporučení léčby, právníkům pro klasifikaci konkrétních situací podle právních norem, železničářům pro hodnocení opotřebení vozů a kolejí.
- V medicíně umožňuje rozpoznávání obrazů identifikovat a klasifikovat orgány pro plánování chirurgických zákroků.
- Online obchody využívají mechanismy strojového učení pro přesnější doporučení produktů pravidelným zákazníkům.

- Autonomní robotizovaní průvodci v muzeích jsou schopni provádět exkurze a odpovídat návštěvníkům na otázky týkající se výstavy.
- V současné vojenské sféře již můžeme vidět počátky autonomního rozhodování: systémy blízké protivzdušné a protiraketové obrany vylučují člověka z rozhodovacího řetězce kvůli nízké rychlosti reakce lidských operátorů, a moderní protilodní střely jsou schopny distribuovat rakety v salvě mezi cíle v závislosti na jejich důležitosti.

### **Perspektivy využití UI**

Očekává se, že v průběhu tohoto desetiletí bude UI schopna řešit ještě více úloh. Níže jsou uvedeny některé z nich, a orientační pokrok prací v oblasti výzkumu a vývoje.

### **Autonomní dopravní prostředky**

Nejsložitějším úkolem pro autonomní dopravní prostředky je účast v silničním provozu z hlediska omezenému prostoru pro manévrování a množství možných situací, které na silnici vznikají. V roce 2018 je nejvyspělejším řešením na trhu autopilot Audi A8, který je schopen autonomně se pohybovat v podmínkách dopravních zácp na dálnicích.

V březnu 2017 zástupci BMW slíbili, že do roku 2021 představí automobil schopný převážet cestující do místa určení bez zásahu řidiče.

### **Vojenské využití**

Nejperspektivnějším směrem pro vojenskou sféru je možnost pro robotizované platformy automaticky rozpoznat a sledovat cíle, a následně autonomně rozhodnout o jejich zničení. Na konci roku 2017 se v zemích "prvního světa" prováděly řady vědeckovýzkumných prací zaměřené na pozemní, námořní a podmořské komplexy.

### **Robotizované chirurgické procedury**

Očekává se, že chirurgie, včetně mikrochirurgie, která vyžaduje přesný zásah do těla pacienta, bude v blízké budoucnosti automatizována. V roce 2016 byl představen jako demonstrátor technologií robot STAR, schopný provádět operace na měkkých tkáních. S pokroky v rozpoznávání obrazů lze očekávat robotizaci všech chirurgických procedur



odstraňování tkání, díky čemuž bude chirurgie dostupnější a sníží tak zátěž na léčebný personál.

### **Neřešené problémy**

I pro zdánlivě všemocné počítačové algoritmy existují neřešitelné a špatně řešitelné problémy. V současné době je ze zjevných důvodů obtížné vyřešit úkoly, které se těžko formalizují, jako například „napsat román“ nebo „vybrat nejkrásnější fotografii“. Ani matematicky přesná formulace nezaručí úspěšné řešení. Důvodem může být složitost matematického modelování problémů nižší úrovně (například modelování pohybu při učení robota hrát fotbal), složitost samotného problému (například pro logický vývod a dokazování matematických tvrzení neexistují algoritmy výrazně lepší než úplný výčet možných logických řetězců), obrovské množství parametrů a nepřesnost pozorovaného světa (například při hře ve fotbale), a nedostatečná výpočetní síla ve srovnání s lidským mozkiem. Koneckonců, algoritmičká simulace vzájemné interakce  $1 \times 10^{11}$  lidských neuronů je nelehká; K začátku roku 2018 byla největším úspěchem v této oblasti simulace  $1,7 \times 10^9$  neuronů se 2400násobným zpomalením, která proběhla v roce 2013; není přitom jisté, zda simulace potřebného množství mozkových buněk umožní vznik mozkové činnosti v počítačovém modelu.

Zvláště je třeba zmínit nedostatky strojového učení. Strojové učení obvykle předpokládá existenci již klasifikovaných tréninkových dat, ve kterých počítačový algoritmus nachází zákonitosti. Kvůli nedostatečným datům mohou v práci algoritmu nastat situace, kdy vstupní data nepatří do žádné z trénovaných tříd; rozpoznání nového fenoménu na vstupu a automatické vytvoření nové třídy objektů je výzvou; tento úkol lze snadno zkomplikovat přidáním aktivního učení během práce klasifikátoru a časové změny rozpoznávaných tříd. Druhým významným nedostatkem strojového učení je vysoká citlivost: například rozpoznání obličeje lze „oblafnout“ nošením na první pohled neškodných brýlí. V některých případech lze pomocí změn, které jsou pro člověka neviditelné, dosáhnout nesprávné klasifikace fotografie: po zdánlivě nepodstatné manipulaci tak může být panda na fotografii klasifikována jako opice.

Zatímco počítače úspěšně řeší „složitě“ úkoly jako jsou symbolické a číselné výpočty a poráží šachové velmistry, v seznamu nevyřešených problémů je mnoho úkolů, jejichž

odlišnost by se dala formulovat poměrně jednoduše: klasifikace „neznámých“ obrazů bez tréninku na předem klasifikovaných vzorcích (například rozpoznávání jablek na fotografii, pokud jsou známy pouze třídy „třešně“ a „hrušky“), motorika, rozumové uvažování. V literatuře je tento fenomén známý jako „Moravcův paradox“. V mnohém to odráží lidské vnímání: schopnosti, které se vyvíjely miliony let evolucí a jsou přítomné u každého dospělého člověka, se zdají být samozřejmé, zatímco matematické problémy jako nalezení nejkratší cesty na mapě se zdají být nepřírozené a jejich řešení nejednoznačné.

### **Nebezpečí umělé inteligence**

V apokalyptických scénářích vědeckofantastických děl se obvykle umělá inteligence v nějakém okamžiku rozhodne zničit lidstvo a vědci tomu nejsou schopni zabránit, neboť si této hrozby nejsou vědomi. Ve skutečnosti je otázka nebezpečí umělé inteligence diskutována ve vědecké komunitě již docela dlouho. Za hlavní nebezpečí je považováno nesprávné cílení „silné“ UI, která disponuje významnými výpočetními a materiálními zdroji a nebere v úvahu zájmy lidí. Výše citovaná kniha Bostroma nabízí možná řešení tohoto problému.

### **Reálnost vzniku silné umělé inteligence**

Přehnaná nebo podceňovaná očekávání vědeckotechnického pokroku vedou k předpovědím, ve kterých bude silná umělá inteligence k dispozici již zítra (nejpozději příští rok), nebo je v zásadě nedosažitelná. Ve skutečnosti je vytvoření silné umělé inteligence obtížné předpovědět, jelikož závisí na úspěšném řešení několika inženýrských a matematických problémů neznámé složitosti. Historický příklad řešení Fermatovy věty, který se objevil 350 let po její formulaci, neumožňuje s jistotou předpovídat termíny řešení problémů takového rozsahu.

### **Mezinárodní tendence**

Rozšíření hranic použitelnosti umělé inteligence vedlo ke zvýšenému zájmu vojenských a s nimi souvisejících kruhů o možnostech autonomních systémů. Kromě výzkumu a vývoje ve výše zmíněných tématech propukla v mezinárodním společenství diskuse o možném omezení nebo dokonce zákazu robotických komplexů. Nejznámější kampaní je "Stop Killer Robots", která aktivně požaduje úplný zákaz vývoje autonomních bojových systémů z etických důvodů. V této souvislosti stojí nicméně za zmínku nejen bojové roboty, ale i

klasifikační systémy, které mají vliv na rozhodnutí o použití síly výhradně na základě metadat, aniž by braly v úvahu výpověď podezřelých.

Veřejná pozornost věnovaná autonomním bojovým systémům vedla k jednáním v rámci *Konvence OSN o zákazu nebo omezení používání určitých druhů konvenčních zbraní, které mohou být považovány za způsobující nadměrné škody nebo mající neselektivní účinek*, pod kterou spadají protipěchotní miny a oslepující lasery. Nicméně začátkem roku 2018 nevedl diplomatický proces k žádným vzájemným závazkům. To je do značné míry způsobeno složitostí v definování "autonomního bojového systému", existencí komplexů (především protivzdušné obrany a protiraketové obrany), odpovídající možné definici, a neochotou vzdát se perspektivní technologie.

Omezit důvody pro kontrolu nad autonomními zbraněmi pouze na morálně-etické aspekty by však bylo přílišným zjednodušením. V jejich prospěch hovoří také otázka strategické stability: za první, použití autonomních systémů může vést k nekontrolovatelné eskalaci vojenských konfliktů s nepředvídatelnými důsledky; za druhé, je obtížné kontrolovat export používaných technologií: pokud je export rakety s plochou dráhou letu viditelný a její dosah je omezen fyzickými parametry, kontrolovat softwarový kód na takové úrovni není možné: oblast použití algoritmů není omezena pouze na vojenskou sféru, a omezení výzkumu až zákaz publikací v oblasti s tak širokým spektrem použití nevyhnutelně narazí na odpor vědecké společnosti, která je do značné míry závislá na mezinárodní spolupráci a velkých konferencích.

## **Umělá inteligence – prospěch nebo hrozba pro lidstvo?**

Kolonin A.G.

### **Hlavní problémy vývoje a potenciální hrozby umělé inteligence**

Navzdory mnoha současným úspěchům v neurobiologii dosud nikdo přesně neví, jak je uspořádán přirozený intelekt. Podobně také nikdo neví, jak přesně vytvořit umělou inteligenci. Existuje řada známých problémů, které je pro její vytvoření nezbytné vyřešit,

a různé názory na důležitost dosažení těch či oněch řešení. Například Ben Goertzel, vedoucí mezinárodních projektů zaměřených na vytváření umělé inteligence s otevřeným zdrojovým kódem OpenCog a SingularityNET, věří, že všechny nezbytné technologie pro vytvoření obecné UI jsou v zásadě již vyvinuty, stačí je pouze nějak správně spojit, aby z toho vyplynula synergie, která povede ke vzniku obecné UI. Jiní odborníci jsou více skeptičtí a domnívají se, že je nutné principiální řešení mnoha níže uvedených problémů. Také se velmi liší expertní odhady časového horizontu vzniku silné UI – od desítky let až po několik desítek let.

Vznik pouze autonomních nebo adaptivních systémů, natož obecných nebo silných UI, je spojen s několika hrozbami různého měřítka, které jsou aktuální již dnes. Patří mezi ně:

- vytvoření takzvaných „smrtících autonomních zbraňových systémů“ – Lethal Autonomous Weapons Systems (LAWS), například dronů pro nájemné vraždy;
- nové kolo závodu ve zbrojení, kdy budou státy zdokonalovat úroveň inteligence autonomních prostředků ničení;
- inteligentní vojenský, průmyslový, nebo domácí systém je schopen nejen cílené akce, ale i vědomého cílení, přičemž autonomní stanovení cílů systémem může vést k nastavení cílů, které jsou v rozporu s cíli člověka;
- vyloučení drtivé většiny populace z výrobního procesu kvůli ještě větší automatizaci, což může vést k ještě větší sociální stratifikaci, snížení efektivity „sociální mobility“ a zvýšení počtu „nadbytečných lidí“ s příslušnými sociálními důsledky.

Náročnost kontroly nad systémy umělé inteligence je dnes způsobena, mimo jiné, „omezeností“ existujících aplikovaných řešení na základě „hlubokých neuronových sítí“, které neumožňují ověřit jak správnost přijímání rozhodnutí před jejich provedením, tak ani zpětně analyzovat rozhodnutí, které bylo strojem přijato. Řešením této problematiky se v současné době zabývá nový směr "vysvětlitelná umělá inteligence" (Explainable Artificial Intelligence, EAI). Pozornost je také věnována snaze o integraci asociativní neuronové sítě a symbolického (založeného na logice) přístupu k problému.

V srpnu 2018 v Praze na půdě ČVUT (České vysoké učení technické) proběhly současně konference zaměřené na umělou inteligenci lidské úrovně, obecnou umělou inteligenci, biologicky inspirovaným kognitivním strukturám (BICA), a také neuro-symbolickým

technologíím. Na konferencích byly prezentovány příspěvky předních specialistů společností a institucí v oblasti výzkumu umělé inteligence (UI): Microsoft, Facebook, DARPA, MIT, Good AI. Během výstupů byl nastíněn jak současný stav vývoje v oblasti UI a nevyřešené problémy stojící před společností, tak i hrozby vznikající v průběhu dalšího rozvoje této technologie. V tomto přehledu se pokusím stručně označit hlavní problémy a hrozby, stejně jako možné způsoby čelit těmto hrozbám.

Nicméně, především je nutné upřesnit význam některých termínů, které jsou obvykle používány společně s UI v různých kontextech: slabá nebo specializovaná UI, autonomní UI (Autonomous AI), adaptivní UI (Adaptive AI), obecná UI (Artificial General Intelligence, AGI), silná UI (Strong AI), UI lidské úrovně (Human-Level AI), UI nadlidské úrovně (Super-human AI).

**Slabá nebo specializovaná UI** obsahuje všechna možná existujícími řešeními jedné konkrétní úlohy a je založena na schopnosti automatizace řešení, ať už jde o hru v Go nebo rozpoznávání tváří na kamerách. Nemá však možnost samostatného učení se jiným úkolům bez přeprogramování člověkem.

**Autonomní UI** je založena na schopnosti systému fungovat dlouhou dobu bez účasti operátora. Umožňuje například dronu se solárními panely uskutečnit několikadenní cestu z Elysejských polí na Rudé náměstí nebo v zpět, zatímco samostatně volí trasu i místa pro mezipřistání k dobití baterií, přičemž se vyhýbá všem možným překážkám.

**Adaptivní UI** je schopna přizpůsobit se novým podmínkám, získávat znalosti, které nebyly naprogramovány při jeho tvorbě. Umožňuje například ruskojazyčnému dialogovému systému na základě studia učebních materiálů nebo při vystavení novému jazykovému systému samostatně si osvojit nové jazyky a uplatnit jejich znalost v rozhovoru.

**Obecná UI** je natolik adaptabilní, že může být za předpokladu patřičného zaučení použita v systémech při nejrůznějších druzích činnosti. Zaučení může být jak samostatné, tak cílené (s pomocí instruktora). V tomto smyslu se ve srovnání se slabou nebo specializovanou UI často používán termín silná UI.

**UI lidské úrovně** je úrovní adaptability srovnatelná s lidskou, tj. systém je schopen osvojit si za stejný časový úsek stejné dovednosti, jako člověk.

**UI nadlidské úrovně** je schopna ještě vyšší adaptability a rychlosti učení. Systém si tak může osvojit znalosti a schopnosti, které jsou pro člověka prakticky nemožné.

### **Principiální problémy vytvoření silné umělé inteligence**

Vznik silné umělé inteligence je zcela logickým důsledkem celkového evolučního procesu, stejně jako vznik molekul z atomů, buněk z molekul, organismů z buněk, diferenciací specializovaných buněk do centrálního nervového systému, vznik sociálních struktur, rozvoj řeči, písma a v konečném důsledku informačních technologií. Zvyšující se složitost informačních struktur a způsoby organizace v průběhu evoluce jsou dobře popsány Valentinem Turchinem. Pokud nedojde k zániku lidské civilizace, bude taková evoluce nevyhnutelná, a v nejděším časovém horizontu bude znamenat záchranu lidstva, jelikož přežití nevyhnutelného zániku Sluneční soustavy a uchování informačního kódu naší civilizace ve Vesmíru bude možno pouze nebiologickou formou existence.

Je důležité si uvědomit, že pro vytvoření silné umělé inteligence není nezbytné chápat fungování přirozeného intelektu, stejně jako není nutné chápat, jak létá pták, aby bylo možné vytvořit raketu. Tak či onak k tomu s největší pravděpodobností k tomu dříve nebo později dojde.

Mezi zásadní problémy, jejichž řešení je pro vytvoření obecné nebo silné UI nezbytné, většina expertů vyzdvihuje následující:

- **Rychlé učení** (few-shot learning) – nutnost vytvoření systémů, které se učí z malého množství dat, na rozdíl od stávajících systémů hlubokého učení, vyžadujících velké objemy speciálně připraveného učebního materiálu.

- **Silná generalizace** (strong generalisation) – vytvoření technologií rozpoznávání situací, ve kterých se rozpoznávané objekty vyskytují v podmínkách odlišných od těch, ve kterých se vyskytovaly v učebním materiálu.

- **Generativní nebo generující modely učení** (generative models) – vývoj technologií učení, jejichž objektem zapamatování jsou, namísto vlastnosti objektu, principy jeho formování, což může prohloubit pochopení podstaty rozpoznávaných objektů a umožnit tak rychlejší učení a silnější generalizaci

- **Strukturované učení a predikce** (structured prediction and learning) – rozvoj technologií, založených na učení formou vícevrstevných hierarchických struktur, kde prvky nižší úrovně definují ty vyšší, což by mohlo být alternativním řešením problémů rychlého učení a silné generalizace.

- **Řešení problému katastrofického zapomínání** – typický problém, který postihuje většinu stávajících systémů, jež jsou nejprve vyučeny rozpoznávat určitý typ objektů a následně jsou upraveny k rozpoznání nového typu objektů, čímž ztrácejí schopnost rozpoznat objekty původního typu.

- **Dosáhnutí možnosti inkrementálního učení** – schopnost systému akumulovat znalosti, postupně si zdokonalovat své možnosti a získávat nové znalosti vzhledem k dialogovým systémům v přirozeném jazyce, aniž by přitom ztratil dříve získané znalosti. Ideálem je splnění "dětského Turingova testu", v rámci kterého musí systém prokázat schopnost postupně si osvojit jazyk od úrovně kojence až po úroveň dospělého člověka.

- **Řešení problému vědomí** – vytvoření ověřeného pracovního modelu vědomého chování, které je schopno efektivní predikce a cíleného chování na základě "vnitřního obrazu světa", v rámci kterého je možné hledat optimální strategie chování pro dosažení stanovených cílů bez přímé interakce s reálným světem. Tímto se významně zvyšuje bezpečnost, ověřování hypotéz, a zároveň rychlost a energetickou efektivitu tohoto ověřování, čímž vytváří možnost autonomního učení (self-learning) živého nebo umělého systému ve "virtuálním světě" vlastního vědomí. Z hlediska aplikace má problém vědomí dvě stránky. Vznik vědomých systémů z jedné strany zvýší jejich efektivitu, ze strany druhé přinese jak dodatečná rizika, tak i otázky etického charakteru. Takové systémy mohou být totiž úrovní sebeuvědomění považovány za rovné člověku, včetně právních důsledků vyplývajících z jeho porušení.

## **Potenciální hrozby UI**

Vznik systémů, ať už autonomní, adaptivní, nebo dokonce obecné či silné UI, je spojen s několika hrozbami různého rozsahu, které jsou aktuální již dnes.

Za prvé, hrozbu pro člověka může představovat inteligence nejen silná, obecná, na úrovni lidské či nadlidské. Postačí k tomu autonomní systém, operující s velkými objemy informací a s velkými rychlostmi. Takzvané „autonomní systémy smrtících zbraní“ – Lethal Autonomous Weapons Systems (LAWS) mohou být totiž vytvořeny pouze na základě autonomního systému, operujícího velkou rychlostí s velkým objemem dat. Typickým příkladem jsou například drony, tisknuté pomocí 3D tiskáren, které slouží k nájemným vraždám. Ty se vyrábí jak v hromadném měřítku, tak i malými skupinami v domácích podmínkách.

Za druhé, hrozbu pro stát může představovat situace, kdy jiný stát (potenciální nepřítel) získá zbraně s obecnou UI s vyšší adaptivitou a autonomií, vyšší reakční rychlostí a prediktivní schopností.

Za třetí, hrozbu pro celý svět představuje situace vyplývající z předchozí hrozby, kdy státy vstoupí do nového kola závodu ve zbrojení, během kterého budou zdokonalovat úroveň inteligence autonomních zbraňových systémů, tak jak to před několika desítkami let předpověděl Stanisław Lem.

Za čtvrté, hrozbu pro jakoukoli stranu představuje jak vojenský, tak i průmyslový nebo domácí inteligentní systém s určitou mírou autonomie a adaptability, schopný nejenom cíleného jednání, ale i vědomého stanovení cíle. Autonomní stanovení cílů systémem může vést k stanovení cílů, které jsou v rozporu s cíli člověka a lidí, přičemž možnost dosažení těchto cílů u systému bude díky jeho vyšší reakční rychlosti, většímu objemu zpracovávaných informací a lepší prediktivní schopnosti, mnohem více. Bohužel rozsahy právě této hrozby nejsou komunitou plně prozkoumány a pochopeny. Rozsah této hrozby bohužel není společností plně prozkoumán a pochopen.

Za páté, hrozbu pro společnost představuje přechod na novou úroveň výrobních procesů v kapitalistické (nebo totalitní) společnosti, kdy malá část populace získá možnost



kontrolovat materiální výrobu, ze které vyloučí ještě větší automatizací převážnou část populace. To může vést k ještě většímu sociálnímu rozvrstvení, snížení efektivity „sociálních výtahů“ a zvýšení počtu „nadbytečných lidí“, což s sebou ponese odpovídající sociální následky.

Nakonec, hrozbu pro lidstvo jako celek může představovat autonomizace globálních výpočetních systémů zpracování dat, šíření informací a přijímání rozhodnutí na základě globálních sítí. Rychlost šíření informací a rozsah působení takových systémů tak může vést k sociálním jevům, které jsou na základě stávající zkušeností a existujících modelů řízení nepředvídatelné. Například, zaváděný systém sociálního kreditu v současné Číně je unikátním experimentem globálního rozsahu s dnes ještě nejasnými důsledky.

Náročnost kontroly nad systémy umělé inteligence dnes je způsobena, mimo jiné, „omezeností“ existujících aplikovaných řešení na základě „hlubokých neuronových sítí“, které neumožňují ověřit jak správnost přijímání rozhodnutí před jejich provedením, tak ani zpětně analyzovat rozhodnutí, které bylo strojem přijato. Řešením této problematiky se v současné době zabývá nový směr "vysvětlitelná umělá inteligence" (Explainable Artificial Intelligence, EAI). Pozornost je také věnována snaze o integraci asociativní neuronové sítě a symbolického (založeného na logice) přístupu k problému.

## **Podivný nový svět bez práce**

Markotkin N.M.

Co nás napadne, když přemýšlíme o blížícím se všudypřítomném zavádění robotů a technologií umělé inteligence (UI)? Mizející fronty na podporu nezaměstnanosti? Drony "Skynetu" kroužící nad spálenými chatrčemi? Nebo naopak, úplná nečinnost a komunismus, zajištěný prací mechanických otroků? Realita pravděpodobně nebude tak extrémní, což však není důvodem nezamýšlet se nad sociálními důsledky technologických změn, odehrávajících se před našima očima.

## **Revoluce na pochodu**

Svět rychle vstupuje do fáze IV. průmyslové revoluce, jejímiž klíčovými složkami jsou robotizace, rozvoj bio a nanotechnologií, 3D tisku, internetu věcí, genetiky a technologií

umělé inteligence. Nadcházející technologické změny budou mít přímé důsledky pro celou řadu stávajících profesí a zcela transformují trh práce, alespoň v rozvinutých zemích.

Vysoké tempo probíhajících změn (10 z nejžádanějších specializací v roce 2010 neexistovalo v roce 2004) komplikuje predikce jejich dopadu na společnost. V této souvislosti se odhady expertů a mezinárodních organizací pohybují od optimistických po alarmující. Pokud opomeneme radikální scénáře, můžeme přesto s jistotou říci, že nás v dohledné budoucnosti čeká zásadní restrukturalizace globální ekonomiky, srovnatelná s I. průmyslovou revolucí XVIII. –XIX. století.

Podle zprávy „Budoucnost povolání“ Světového ekonomického fóra (WEF), 65 % současných žáků základních škol bude mít zcela nové, zatím neexistující profese. K podobným závěrům dospěla společnost McKinsey, která ve své zprávě zdůrazňuje, že při současné úrovni technologického vývoje může být 30 % funkcí obsažených v 60% povolání automatizováno. M. Osborne a K. Frey z Oxfordské univerzity přináší ještě pesimističtější prognózu. Podle výsledků jejich studie bude v příštích 20 letech v USA automatizací ohroženo 47 % pracovních míst.

### **Které profese budou nahrazeny**

Které profese jsou tedy ohroženy? Především jde samozřejmě o nekvalifikovanou rutinní práci. Ve studii M. Osborna a K. Freye patří mezi nejohroženější: úředníci, zadavatelé dat, knihovníci, obsluha strojů, instalatéři, prodejci, technici nastavující zařízení a další.

Podle odhadů Světového ekonomického fóra dojde mezi lety 2015 a 2020 k největšímu snížení pracovních míst v kancelářských profesích (4,91 %) a ve výrobním sektoru (1,63 %). Zhruba o 1 % by měla klesnout také zaměstnanost v oblastech jako design, zábavní průmysl, stavebnictví a prodej. Naopak největší nárůst pracovních míst je předpokládán v oblasti počítačových technologií (3,21 %), architektonických a inženýrských specializací (2,71 %) a managementu (o necelé 1 %).

Ve střednědobém horizontu budou pravděpodobně automatizací ohroženy také profese spojené s dopravou. Rozvoj technologie autonomních vozidel může radikálně změnit trh jak osobní, tak zejména nákladní dopravy. Pouze v USA je v odvětví dálkové nákladní přepravy zaměstnáno 8,7 milionu lidí. Pokud vezmeme v úvahu celý byznys spojený

s řidiči kamionů (motely, restaurace u cesty atd.), může toto číslo vzrůst až na 15 milionů, což činí přibližně

10 % práceschopné populace země. Neméně důležitým může být snížení pracovních míst i v sektoru veřejné osobní dopravy. Existuje vysoká pravděpodobnost, že autonomní technologie budou v blízké budoucnosti zavedeny také v námořní nákladní přepravě. S rozvojem technologií umělé inteligence čekají těžké časy také právníky, učitele, horníky, střední management, novináře a celou řadu dalších profesí.

Hlavní zaměstnanost se tak postupně přesune ze sektoru služeb do jiných sektorů ekonomiky, z nichž mnohé mají být teprve vytvořeny. Tato perspektiva však není unikátní. Potvrzuje tím pouze revoluční charakter probíhajících změn. Před I. průmyslovou revolucí bylo více než 70 % populace zaměstnáno v zemědělství, zatímco dnes je tento ukazatel v rozvinutých zemích na úrovni několika procent. Podíl zaměstnanosti v průmyslu rostl až do poloviny 20. století, v důsledku digitální revoluce však klesl na současných 24 % v EU a 19 % v USA (v Rusku tento ukazatel činí 27 %). Přestože počet pracovníků klesá, objem výroby stále stabilně roste. Nyní je zdá se, na řadě, automatizace služeb.

### **Zlatý věk inženýrů a psychiatrů?**

Nejméně v krátkodobém horizontu utrpí profese vyžadující intelektuální úsilí, nebo přímý osobní kontakt s klientem. Podle studie Oxfordské univerzity jsou automatizací nejméně ohroženy různé lékařské a psychologické specializace, trenéři, sociální pracovníci, programátoři, inženýři, vyšší management a kreativní profese. Jinými slovy, nejlépe připraveni jsou ti, jejichž práce vyžaduje kreativní přístup a nezakládá se na vykonávání určitých naučených kombinací. U inženýrů je důležité upřesnit, že v "bezpečí" jsou především projektoví inženýři, zatímco provozní inženýři spadají naopak do rizikové oblasti.

Automatizaci kreativních profesí brzdí tři klíčové faktory. Pro úspěšné plnění jejich úkolů musí umělá inteligence disponovat vnímáním, možností manipulace s materiálními objekty (hmatem), kreativním intelektem a sociálním intelektem. Současná úroveň technologického vývoje však tyto problémy vyřešit neumožňuje. Nicméně s rozvojem technologií "silné" UI bude spektrum prací, které jsou jí dostupné, neustále růst. Stávajícím technologiím tak rozšíří hranice automatizace a dá počítačům možnost přijímat řídicí rozhodnutí, nebo se dokonce věnovat kreativní činnosti. Proto nelze vyloučit, že ve

střednědobém nebo dlouhodobém horizontu stroje úspěšně nahradí nejen inženýry a manažery, ale i spisovatele a umělce. Zvláště když existují úspěšné případy vytvoření uměleckého textu umělou inteligencí.

Je tak zcela možné, že většina práce schopné populace se v dohledné budoucnosti znovu posadí do školních lavic. Problém je však ten, že nikdo přesně neví, co přesně se bude muset učit. Podle některých odhadů až 85 % profesí potřebných v roce 2030 dnes ještě neexistuje. Nové skutečnosti se současné vzdělávací systémy ještě nepřizpůsobily, a to ani v rozvinutých zemích.

## **Etické a právní otázky umělé inteligence**

Karljuk M.V.

Etika a právo jsou v moderní společnosti úzce spojeny a mnoho právních rozhodnutí vychází z vnímání určitých etických problémů. Umělá inteligence dodává těmto otázkám nový rozměr. Systémy využívající umělou inteligenci jsou z hlediska složitosti úkolů, které mohou vykonávat, a potenciálního vlivu na svět, stále samostatnější, a schopnost člověka porozumět, předvídat a kontrolovat jejich fungování se snižuje. Většina lidí skutečnou autonomii takových systémů podceňuje. Mohou se učit z vlastní zkušenosti a provádět akce, které jejich tvůrci nezamýšleli. To vede k řadě etických a právních problémů, které budou v tomto článku rozebírány.

### **Etika a umělá inteligence**

Existuje poměrně známý experiment, nazývaný tramvajové dilema. Otevírá řadu důležitých etických otázek, které přímo souvisí s umělou inteligencí. Představte si, že se tramvaj nekontrolovatelně řítí po kolejích, k nimž je přivázáno pět lidí. Stojíte u výhybky, pomocí které můžete změnit směr a tramvaj pojedje jinou cestou, na které je ke kolejím přivázán jeden člověk. Přepnete výhybku? Na tuto otázku neexistuje jednoznačná odpověď. Navíc existuje mnoho variací situace, ve kterých je nutné podobné rozhodnutí učinit. Různé sociální skupiny tak poskytují různé odpovědi. Budhističtí mniši jsou například ochotni obětovat život jednoho člověka, aby jich zachránili pět, a to i ve složitější variantě tramvajového dilematu.

Pokud jde o umělou inteligenci, taková situace může nastat například na silnici, po které se pohybuje autonomní vozidlo. V případě, že je nehoda nevyhnutelná, vzniká tak otázka: či život má být prioritou – cestujících, chodců nebo ani jednoho?

Na Massachusettském technologickém institutu byla dokonce vytvořena speciální webová stránka věnovaná této problematice, kde mohou uživatelé vyzkoušet různé scénáře a rozhodnout, jak by se mělo v dané situaci jednat.

Nastávají také otázky, co je povoleno z právního hlediska? Na základě čeho dělat výběr? A kdo ponese vinu? Na tento problém již reagovaly firmy i regulátoři. Zástupci firmy "Mercedes" například přímo řekli, že jejich vozy budou dávat přednost pasažérům, na což německé ministerstvo dopravy ihned reagovalo odpovědí, predikující příchozí regulaci – že dělat takový výběr na základě řady kritérií bude nezákonné a výrobce ponese v každém případě odpovědnost.

Jiné země mohou zvolit jinou cestu. Vezměme si například čínský systém sociálního skóre, podle kterého je každému občanovi přiřazeno hodnocení v závislosti na dodržování zákonů, prospěšné společenské činnosti atd. Osoby s nízkým kreditem budou podléhat sankcím. Co brání čínské vládě přinutit výrobce autonomních vozidel v případě nevyhnutelné nehody obětovat osobu s nejnižším skóre? Díky technologiím rozpoznávání obličejů a přístupu k příslušným databázím je docela možné identifikovat a porovnat hodnocení potenciálních obětí.

### **Hlavní problémy, kterým čelí právo**

Právní problémy jsou ještě hlubší, a to zejména v případě robotů. Systém, který se učí na základě informací získaných z vnějšího světa, může jednat způsobem, který tvůrci nemohli předvídat. Předvídatelnost je však pro současné právní přístupy klíčová. Takové systémy mohou navíc jednat nezávisle na svých tvůrcích nebo operátorech, což komplikuje určení subjektu odpovědnosti. Vystává tak problém předvídatelnosti a schopnosti jednat samostatně, ale bez právní odpovědnosti.

Možností regulace je mnoho. Regulacemi, založenými na současných normách lze například regulovat technologie využívající umělou inteligenci jako předměty autorských práv nebo jako majetek. Komplikace však nastávají v případě autonomního jednání mimo vůli tvůrce, vlastníka nebo držitele. V této souvislosti lze aplikovat normy, které regulují

specifický druh majetku, konkrétně zvířata, která jsou také schopna autonomního jednání. V ruském právu se v takovém případě aplikují obecná pravidla o majetku (čl. 137 Občanského zákoníku RF), a otázka odpovědnosti tak bude řešena podle čl. 1064 Občanského zákoníku RF: škoda způsobená osobě nebo majetku občana podléhá plné kompenzaci osobou, která škodu způsobila, čímž je vlastník autonomního agenta.

Návrhy na využití právního postupu aplikovaného u zvířat byly předloženy, mají však svá omezení. Za prvé, analogické použití legislativy je v rámci trestního práva nepřipustné. Za druhé, tyto normy byly vytvořeny především pro domácí zvířata, která za běžných okolností škodu nezpůsobují. U pokročilejších systémů existuje možnost provést analogii s divokými zvířaty, pro která existují přísnější pravidla. Ale i zde vzniká otázka, jak provést takové rozdělení v souvislosti s charakteristikami umělé inteligence, uvedenými výše. Přísná pravidla mohou navíc kvůli velkým a nepředvídatelným rizikům odpovědnosti tvůrců vést ke zpomalení zavádění technologií umělé inteligence.

Častým návrhem je uplatnit na takové systémy zákon o právnických osobách. Jelikož je právnická osoba uměle vytvořeným subjektem práva, podobným statusem by mohly být obdařeny i roboty. Právo je dostatečně flexibilní, a může tak přiznat nebo omezit práva prakticky komukoli. Například otroci v minulosti prakticky žádná práva neměli a byli považováni za majetek. Existuje také opačná situace, kdy jsou právy obdařeny objekty bez schopnosti jednat. I dnes existují příklady neobvyklých objektů, uznávaných jako právnické osoby, a to jak ve vyspělých, tak i v rozvojových zemích. Na Novém Zélandu byl v roce 2017 přijat zákon, který uznal status právnické osoby řece Whanganui. Zákon uvádí, že daná řeka je právnickou osobou a má všechna práva, zmocnění a povinnosti právnické osoby. Tímto zákonem byla řeka transformována z majetku na právnickou osobu, což rozšířilo hranice chápání toho, co může být považováno za majetek a co ne. V roce 2000 uznal Nejvyšší soud Indie status právnické osoby posvátnému textu sikhů, Guru Granth Sahib.

Nehledě na tyto extrémní případy, běžné společnosti nesou jakožto právnické osoby v rámci právního systému občanskoprávní, někdy i trestněprávní odpovědnost. Nehledě na to, že společnosti (nebo státy) nedisponují svobodnou vůlí, záměrem, nebo schopností úmyslně a vědomě jednat, mohou být za určité činy právně odpovědné. Stejným způsobem mohou být za svoje jednání odpovědné roboty, aniž by měli svobodnou vůli, nebo jednali záměrně.

Analogie s právními osobami je nicméně problematická. Koncept právnické osoby sice slouží pro rychlé a efektivní vykonávání spravedlnosti, činy právnických osob však vždy vyplývají z činů jednotlivce nebo skupiny lidí, i když je nelze přesně určit. Jinými slovy, právní odpovědnost společností a podobných entit je spojena s činy jejich zástupců nebo zaměstnanců. Trestní odpovědnost právnické osoby je navíc možná pouze v případě, pokud dojde k identifikaci fyzické osoby, která trestný čin jménem právnické osoby spáchala. Činy systémů umělé inteligence nebudou vždy přímo vyplývat z činů člověka.

U takových systémů lze uplatnit právní normu objektivní odpovědnosti. V souladu s odst. 1 čl. 1079 Občanského zákoníku Ruské federace jsou právnické osoby a občané, jejichž činnost je spojena se zvýšeným nebezpečím pro okolí (používání dopravních prostředků, mechanismů atd.), povinni nahradit škodu způsobenou zdrojem zvýšeného nebezpečí, pokud nedokážou, že škoda vznikla v důsledku vyšší moci nebo úmyslu poškozeného. Problém spočívá v rozpoznání, které systémy by mohly představovat zvýšené riziko. Tato otázka je podobná výše uvedenému problému rozlišení domácích a divokých zvířat.

## **Národní a mezinárodní regulace**

Právní podmínky pro rozvoj technologií využívajících umělou inteligenci aktivně vytváří mnoho zemí. V Jižní Koreji například od roku 2008 existuje zákon „O podpoře vývoje a šíření inteligentních robotů“. Zákon má za účel zlepšení kvality života a rozvoj ekonomiky prostřednictvím vývoje a propagace strategie udržitelného rozvoje průmyslu inteligentních robotů, přičemž každých pět let vypracovává vláda hlavní plán pro efektivní dosažení tohoto cíle.

Zde uvedeme dva nedávné příklady – Francii, která projevila ambice stát se evropským a světovým lídrem v oblasti umělé inteligence; a Evropskou unií, kde byly navrženy pokročilé normy regulace inteligentních robotů.

## **Francie**

Na konci března 2018 prezentoval francouzský prezident E. Macron národní strategii v oblasti umělé inteligence. Francie plánuje v průběhu pěti let investovat 1,5 miliardy eur na podporu výzkumu a inovací v této oblasti. Strategie vychází z doporučení, obsažených v reportu, vypracovaném pod vedením francouzského matematika a poslance Národního shromáždění Francie, Cedrica Villaniho. Strategie se tak zaměří na čtyři konkrétní sektory: zdravotnictví, dopravu, životní prostředí a obranu a bezpečnost. Cílem je tedy využít potenciálu komparativních výhod a odborných znalostí se zaměřením na sektory, důležité z hlediska veřejného zájmu, a ve kterých budou společnosti moci hrát klíčovou roli na globální úrovni.

Celkem bylo předloženo sedm klíčových návrhů, z nichž jeden je pro účely tohoto článku zvláště zajímavý – udělat umělou inteligenci dostupnější. Algoritmy jsou uzavřené a ve většině případů jsou obchodním tajemstvím. Nicméně algoritmy mohou být zaujaté. Během autonomního učení mohou například absorbovat stereotypy existující ve společnosti nebo předané vývojáři a na jejich základě přijímat rozhodnutí. Na tomto principu jsou již přijímána soudní rozhodnutí. V USA byl obžalovaný odsouzen k dlouhodobému trestu odnětí svobody na základě informací získaných od algoritmu, který hodnotil možnost opakování trestné činnosti. Obžalovaný neúspěšně napadal použití algoritmu k takovému rozhodnutí, protože nebyla poskytnuta hodnotící kritéria, která jsou považována za obchodní tajemství. Francouzská strategie navrhuje rozvinout transparentnost algoritmů a možnosti jejich ověření, definovat etickou odpovědnost těch, kteří pracují v oblasti umělé inteligence, vytvořit etické poradní výbory atd.

## **Evropská unie**

V EU byla prvním krokem směrem k regulaci umělé inteligence Rezoluce Evropského parlamentu z roku 2017, " Občanskoprávní pravidla pro robotiku ". Již v roce 2015 byla v Evropském parlamentu vytvořena pracovní skupina zaměřená na právní otázky spojené s rozvojem robotiky a umělé inteligence v EU. Rezoluce není závazná, nicméně poskytuje řadu doporučení Evropské komise na možná opatření v tomto směru, jak v rámci občanského práva, tak v rámci etických aspektů robotiky.



Rezoluce definuje "inteligentního robota" jako robota, schopného autonomie prostřednictvím použití senzorů a/nebo interakce s okolním prostředím; zároveň má robot alespoň minimální fyzickou podporu, adaptuje své chování a akce v souladu s podmínkami prostředí, přičemž z biologického hlediska není živý. Byl předložen návrh vytvoření registračního systému pro pokročilé roboty, který by byl řízen agenturou EU zaměřenou na robotiku a umělou inteligenci. Tato agentura by rovněž poskytovala technickou, etickou a regulační expertizu v oblasti robotiky. Pokud jde o odpovědnost, byly předloženy dvě možnosti: objektivní odpovědnost (nevyžadující vinu), nebo přístup managementu rizik (odpovědnost osoby, která mohla rizika minimalizovat). Odpovědnost by měla být proporcionální skutečné úrovni pokynů, které jsou robotovi dávány, a úrovni jeho autonomie. Pravidla odpovědnosti mohou být doplněna povinným pojištěním pro uživatele robotů a kompenzačním fondem pro výplatu odškodnění v případě absence pojistné smlouvy, která by pokrývala riziko. Pro řešení etických otázek rezoluce navrhuje dva etické kodexy: Etický kodex pro vývojáře robotiky a Etický kodex pro vědecký výzkum: Ten první uvádí čtyři etické principy: 1) „dělej dobro“ (roboti musí jednat v zájmu lidí); 2) „neškod“ (roboti nesmí způsobovat škodu člověku); 3) autonomie (interakce člověka s roboty by měla být dobrovolná); 4) spravedlnost (výhody získané z činnosti robotů by měly být spravedlivě rozděleny).

Příklady uvedené v článku mimo jiné ukazují, jak společenské hodnoty ovlivňují postoj k umělé inteligenci a jejímu právnímu uznání. Naše postavení k autonomním systémům, ať už jde o roboty nebo něco jiného, a naše přehodnocení jejich role ve společnosti a postavení mezi námi může mít transformační efekt. Právní subjektivita definuje, co je pro společnost důležité, a umožňuje rozhodnout, zda je „něco“ cenným a vhodným objektem pro držení práv a povinností.

V souvislosti s charakteristikami umělé inteligence se objevují návrhy na přímou odpovědnost určitých systémů. Podle těchto názorů neexistují zásadní důvody, proč by autonomní systémy neměly nést odpovědnost za své činy. Otázkou však zůstává, zda je nutné nebo žádoucí takovou odpovědnost zavádět (alespoň v této fázi). To je spojeno také s etickými problémy uvedenými v článku. Možná by bylo efektivnější učinit odpovědnými za činy těchto systémů jejich programátory nebo uživatele. To by však mohlo zpomalit inovace. Proto je nutné pokračovat v hledání optimálního řešení.

K tomu je důležité odpovědět na mnohé otázky. Například, jakých cílů se snažíme rozvojem umělé inteligence dosáhnout? Čí zájmy by měly převažovat? Jak efektivní to bude? Odpověď na tyto otázky pomůže předejít situacím podobným té, která se stala v Rusku 17. století, kdy bylo zvíře (konkrétně kozel) posláno za spáchané činy do vyhnanství na Sibiř.

## 4 Překladový komentář

Součástí této odborné práce je také překladový komentář, ve kterém se zaměřím na způsoby převodu textu z ruštiny do češtiny a využití překladatelských transformací. Cílem překladu bylo dosáhnout co největší shody originálního textu a textu překladu jak po stránce významové, tak po stránce stylistické.

Během překladu jsem se nejčastěji potýkal se složitou terminologií nově se rozvíjejícího technologického odvětví, s chybějícím ekvivalentním lexikem a složitými dlouhými souvětími. Mezi nejčastěji užívané překladové transformace patří kalkování, transplantace, transliterace a záměna gramatické kategorie.

### 4.1. Transliterace

Během transliterace přepisujeme grafickou podobu slova identicky, jak je psáno ve výchozím jazyce. „(босс – boss, Обама – Obama)“ (Vysloužilová, Machalová 2011: 9). Tuto transformaci jsem použil především při překladu vlastních jmen jak ruských, tak anglických.

Ideální transliterace se snaží dosáhnout přepisu 1:1, kde každý znak výchozího jazyka odpovídá jednomu znaku v jazyce cílovém. V praxi to však není vždy možné, protože různá písma mají odlišný počet znaků. Proto se používají diakritická znaménka nebo spřežky (kombinace dvou znaků), aby se zachoval co nejpřesnější přepis (Špačková 2016: 52).

- *Елисейских – Elysejských*
- *Маркоткин Н.М.- Markotkin N.M.*
- *Карлюк М.В.- Karljuk M.V.*
- *В исследовании М. Осборна – Ve studii M. Osborna*

## 4.2. Transformace slovosledu

Další poměrně často používanou transformací byla změna slovosledu. Ruský jazyk má na rozdíl o češtiny jiný pořádek slov, proto bylo nezbytné tyto rozdíly překlenout touto transformací. V ruštině se může přívlastek nacházet za nebo před určitým podstatným jménem. (*книга, лежащая на столе / лежащая на столе книга*) (Žváček 1995: 28).

- *Среди профессий, которые с наименьшей вероятностью страдают в краткосрочной перспективе, чаще всего называют те, которые связаны с интеллектуальным трудом или подразумевают непосредственный персональный контакт с клиентом – Неjméně в краткосрочном горизонте utrpí profese vyžadující intelektuální úsilí, nebo přímý osobní kontakt s klientem*

→ V tomto příkladu jsem zároveň použil konkretizaci při překladu slovního spojení „с наименьшей вероятностью страдают“ za „Неjméně в краткосрочном горизонте utrpí“.

- *В качестве принципиальных проблем, решение которых еще предстоит... - Mezi zásadní problémy, jejichž řešení je ...*

→ Zde je v ruštině slovosled, který se v českém jazyce nepoužívá, proto jsem jej nahradil vztažným zájmenem „jejichž“.

- *Слабый или специализированный ИИ представлен всеми без исключения существующими решениями.. - Slabá nebo specializovaná UI obsahuje všechna možná existující řešení...*

→ Zde jsem opět konkretizoval slovní spojení *всеми без исключения* na *всеchna možná existující řešení*. Vyřešil jsem tak problém změněného pořádku slov.

- *... настолько высокую адаптивность, что обладающая им система может быть использована в самых различных видах деятельности при соответствующем обучении... - ...je natolik adaptabilní, že může být za předpokladu patřičného zaučení použita v systémech při nejrozličnějších druzích činnosti*

→ Kromě změny pořádku slov jsem zde také použil slovnědruhovou záměnu podstatného jména „адаптивность“ za přídavné jméno „adaptabilní“.

- *предполагающего способность системы накапливать знания и совершенствовать свои возможности постепенно, не теряя при этом полученные ранее знания, но приобретая новые знания применительно к системам диалогового общения на естественном языке – schopnost systému akumulovat znalosti, postupně si zdokonalovat své možnosti a získávat nové znalosti vzhledem k dialogovým systémům v přirozeném jazyce, aniž by přitom ztratil dříve získané znalosti...*

### 4.3. Univerbizace

Další častou překladovou transformací byla univerbizace. V takových případech se víceslovné pojmenování nahrazuje jednoslovným (Vysloužilová, Machalová 2011: 11).

Tato transformace je často využívána při překladu z ruštiny do češtiny, protože oba jazyky mají rozdílné tendence ve vyjadřování. Ruština má sklon používat delší, více popisná označení, zatímco čeština upřednostňuje kratší a stručnější výrazy.

- *обратить внимание – roznamenat*
- *проводить анализ – analyzovat*
- *приходит нам в голову – ... nás napadne*
- *Уходящие за горизонт – mizející*
- *скорее всего – pravděpodobně*
- *как минимум – alespoň*
- *В первую очередь – především*
- *наиболее уязвим – nejohroženější*
- *в то время как – zatímco*
- *ставит под вопрос – zpochybňuje* → Slovo *zpochybňuje* jsem tu zvolil z toho důvodu, protože dobře vystihuje nejistotu, která vzniká tvorbou nových forem umělé inteligence.

### 4.4. Generalizace

Generalizace je proces, při kterém je specifický termín nahrazen termínem s širším a obecnějším významem. Tento přístup je užitečný, když je potřeba vyjádřit myšlenku méně konkrétně a více univerzálně (Vysloužilová, Machalová 2011: 10).

- *Рабочая тетрадь – Dokument* → Doslovný překlad by znamenal *pracovní sešit*, což se v tomto kontextu nehodí. Slovo *dokument* lépe vystihuje důležitost a odbornost práce.

#### 4.5. Multiverbizace

Zde dochází k přesnému opaku, kdy je při překladu víceslovné pojmenování nahrazeno jednoslovným (Vysloužilová, Machalová 2011: 11).

- *актуальными – které jsou aktuální*
- *обусловлена – je způsobena*
- *Если же учитывать – Pokud vezmete v úvahu*
- *пассажирских перевозок – veřejné osobní dopravy*
- *предстоит создать – mají být vytvořeny*

#### 4.6. Kalkování

„Kalkování spočívá ve vytvoření analogického vlastního jména, které se s výchozím vlastním jménem shoduje strukturálně, lexikálně i sémanticky“, jako například: Бриллиантовая рука – Briliantová ruka (Špačková, 2016: 29).

Kalkování v translatologii znamená, že se slovo nebo slovní spojení rozdělí na jednotlivé morfémy nebo větší jednotky, které se pak přeloží samostatně do cílového jazyka. Kalkování je možné pouze tehdy, pokud má alespoň jeden prvek vlastního jména obecný význam (Špačková 2016: 77).

- *Искусственный интеллект – umělá inteligence*
- *ИИ сверхчеловеческого уровня – UI nadlidské úrovně*
- *Российского совета по международным делам – Ruské rady pro mezinárodní vztahy*
- *Машинное обучение – strojové učení*
- *Глубинное обучение – hluboké učení*
- *Baby Turing Test – dětský Turingův test*
- *биологически-вдохновленным когнитивным архитектурам – biologicky inspirovaným kognitivním architekturám*

#### 4.7. Záměna gramatické kategorie

V této transformaci dochází ke změně jedné gramatické kategorie na jinou. Jde například o záměnu gramatické kategorie čísla, času, nebo vidu (Vysloužilová, Machalová 2011: 10).

Gramatické transformace mění gramatické kategorie, ale ponechávají nezměněný lexikální obsah. Tyto transformace jsou důležité v odborném překladu, protože reflektují rozdíly mezi jazykovými systémy a odlišnosti v organizaci odborného stylu (Pekařová 2019: 23).

- *Несмотря на множество современных достижений...* - *Navzdory mnoha současným úspěchům...* - V tomto případě dochází k záměně akuzativu (4p) na lokál (6p)
- *...используются совместно с ИИ в различных контекстах* – *používány společně s UI v různém kontextu* – Zde se jedná o záměnu čísla množného za jednotné
- *...были представлены...* - *...byly prezentovány...* - Slovo *prezentovány* tu bylo přeloženo z původně dokonavého vidu do vidu nedokonavého
- *Слабый или специализированный ИИ* – *Slabá nebo specializovaná UI* – Zde byl původně mužský rod změněn na rod ženský

## Závěr

Tato diplomová práce se zabývala překladem odborného textu o umělé inteligenci (UI) a jejím vlivu na různé aspekty moderního světa. Práce byla rozdělena do dvou částí – teoretické a praktické, které se vzájemně doplňují a poskytují komplexní přehled o problematice UI v technologické i společenské sféře.

Hlavním cílem této práce bylo vytvořit ekvivalentní překlad originálního textu o umělé inteligenci, který rozšíří povědomí o negativním dopadu umělé inteligence na společnost a budoucí pracovní sféru. Jelikož se touto stinnou stránkou UI zabývá málo odborných textů v českém jazyce, vedlejším cílem této práce bylo také rozšířit možnost získání těchto informací také českému čtenáři. Teoretická část poskytla historický přehled strojového překladu, přičemž se zaměřila na vývoj UI ve strojovém překladu a její aplikace v lékařství a právu. Demonstrovala na několika aktuálních příkladech z těchto oblastí, jak mohou chyby ve strojovém překladu vést k nechtěným následkům, které mohou negativně ovlivnit životy lidí. Rozebrala také etické otázky spojené s UI, jako je ochrana osobních údajů, vlastnictví dat a předsudky ve strojovém překladu. Tyto otázky jsou důležité v kontextu současného technologického vývoje a rychlého růstu UI, k čemuž se právě obrovské množství dat, textů, soukromých informací a osobních údajů, používá. Čím dál vyšší kvalita strojového překladu s sebou také přináší mnohem větší potřebu post-editace, oproti potřebě mechanického překladu textu. Proto jsem se v teoretické části také věnoval základní analýze a druhům post-editace a jaký v nich je rozdíl. Tento narůstající způsob překladu bude čím dál potřebnější.

Praktická část analyzovala překládaný text o UI, poskytla jeho obsahovou, syntaktickou, morfologickou a lexikální analýzu. Samotný originální text zkoumal, jak UI ovlivňuje různé oblasti života, včetně průmyslu, mezinárodních vztahů a společenských struktur. Zabýval se také vlivem UI na pracovní trh a mezinárodní vztahy, poukázal na důležitost regulace a zodpovědného využití této technologie.

Práce ukázala, jak UI ovlivňuje moderní svět a zdůraznila potřebu vyváženého přístupu k této technologii. Z teoretické i praktické analýzy vyplývá, že UI má významný vliv na různé oblasti, ale její využití přináší také rizika, která je třeba řešit a kterým není věnována dostatečná pozornost především ze strany laické veřejnosti. Důležitost regulace a etických otázek je klíčová pro udržitelné využití UI v budoucnu.

Celkově tato práce poskytuje komplexní pohled na UI, její historii, vývoj a aplikace především v oblasti strojového překladu.

## Резюме

Предметом моей магистерской работы является перевод научного текста, касающийся технологии искусственного интеллекта (ИИ), включая комментарий и глоссарий. Я выбрал этот конкретный текст из-за его актуальности и глобального воздействия, которое искусственный интеллект оказывает на различные аспекты жизни, такие как промышленность, технология и коммуникация. Особый интерес для меня представляют воздействия ИИ на международные отношения и социальные структуры. Хотя искусственный интеллект часто представляется преимущественно положительно, его потенциальные негативные эффекты редко обсуждаются, что способствовало моему решению исследовать эту тему и таким образом расширить мое понимание и переводческие навыки в этой области.

В первой части я подробно исследую историю и развитие машинного перевода с его начала в середине 20 века до современных передовых технологий на основе искусственного интеллекта. Эта история начинается с первых примитивных попыток машинного перевода, основанных на системах, основанных на правилах. Эти ранние системы использовали двуязычные словари и вручную написанные правила для перевода, что было трудно и плохо поддерживаемым для разных языков и областей. Тем не менее, эти системы нашли применение, главным образом, в военных целях и привели к первым экспериментам, как тот, который был проведен в 1954 году на Георгианском университете с использованием компьютера IBM.

Другим значимым этапом в истории машинного перевода стало публикация отчета ALPAC в 1966 году, критикующего эффективность и результаты машинного перевода, что привело к значительному сокращению финансирования и интереса к этой области. Однако с приходом нового тысячелетия и развитием двуязычных корпусов начали применяться методы, основанные на корпусах, такие как перевод на основе примеров (EBMT), статистический машинный перевод (SMT) и нейронный машинный перевод (NMT).

EBMT, который появился в середине 80-х, был направлен на поиск похожих пар предложений в двуязычном корпусе, и, несмотря на свои ограничения, обеспечивал качественные результаты для ограниченного диапазона переводов. SMT, разработанный в 1990 году, позволил автоматически изучать правила перевода из большого объема данных, что представляло собой значительный сдвиг



от ручного написания правил к автоматическому обучению. Этот метод стал более популярным и был далее развит с использованием инструментов и привел к развитию фразовых методов SMT.

Последнее десятилетие принесло значительный прогресс в виде нейронного машинного перевода (NMT), который использует кодирующие и декодирующие сети, что позволяет выполнять перевод на основе более глубокого понимания языка. Эти модели не требуют ручного написания правил и учатся семантическим представлениям непосредственно из тренировочных корпусов. NMT значительно улучшил способность машин обрабатывать контекст и тонкие языковые различия, что драматически повысило качество и эффективность переводов.

В целом можно сказать, что машинный перевод значительно развился от своих начал, основанных на простых правилах, до сложных систем, использующих передовые технологии искусственного интеллекта. Этот прогресс не только улучшил качество и скорость переводов, но также расширил возможности его применения в различных областях от веб-сайтов до приложений. Тем не менее остаются вызовы, такие как перевод между языками с различными структурами и сохранение культурных различий, что указывает на то, что несмотря на достигнутый прогресс, все еще есть место для дальнейших исследований и разработок в этой динамично развивающейся области.

В следующей части своей работы я занимаюсь растущим глобальным спросом на переводческие услуги и сопутствующим давлением на переводчиков, что вызывает необходимость мониторинга и оценки их работы. Мониторинг, хотя часто воспринимается как спорная практика, становится все более распространенным инструментом для повышения эффективности и продуктивности в этой области. В этом контексте Европейский союз поддерживает проекты, направленные на мониторинг в различных областях, включая здравоохранение и безопасность, что демонстрирует более широкий контекст применения этих методов.

В своем анализе я также рассматриваю способы, которыми данные собираются от переводчиков во время их работы, включая запись времени, проведенного над задачами, и активности на клавиатуре. Эти методы мониторинга используются не только для оценки продуктивности, но и для расчета вознаграждений и мониторинга прогресса работы. Я критически рассматриваю

этические аспекты этих практик, особенно учитывая, что многие платформы для переводов начинают внедрять мониторинг как стандартную процедуру.

Для более глубокого понимания последствий этих практик я обращался к исследованию „Translation Productivity“, которое рассматривало, как переводчики воспринимают мониторинг своей деятельности. Результаты показали, что хотя некоторые внутренние переводчики считают мониторинг полезным для повышения своей продуктивности, они также воспринимают его как нагрузку, которая может вызывать чувство постоянного давления на выполнение. С другой стороны, внешние переводчики во время исследования выразили смешанные чувства, от первоначального любопытства до опасений по поводу этических последствий этих методов.

Далее я подчеркивал необходимость прозрачности и внедрения механизмов согласия, которые могли бы помочь смягчить негативные последствия мониторинга и одновременно поддерживать справедливые и этичные практики в области перевода.

Также, я занимался машинным переводом и его применением в областях медицины и права, где языковые барьеры представляют значительную проблему. Машинный перевод, представленный в основном современными нейронными сетями, обеспечивает эффективные и плавные переводы, однако также существуют потенциальные риски, которые могут привести к ошибкам с серьезными последствиями. В медицинском контексте ошибки в переводе могут привести к фатальным ситуациям, в то время как в праве они могут вызвать судебные ошибки.

Исследования указывают на недостатки даже продвинутых систем перевода, таких как Google Translate, который, несмотря на свою популярность и широкое использование, подвержен ошибкам при переводе критически важных фраз. Эти ошибки подчеркивают необходимость более глубокого понимания ограничений технологий и большего осведомленности о возможных ограничениях машинного перевода среди общественности.

В медицинском секторе машинный перевод используется для пациентов с ограниченными языковыми способностями, где он может облегчить коммуникацию, но исследования показывают, что даже после обновления алгоритмов он может содержать серьезные ошибки, которые могут угрожать здоровью пациента.

В сфере права недостаточная точность машинного перевода может привести к недоразумениям, которые могут повлиять на исход судебных споров. Примеры из практики показывают, как переводы могут влиять на судебные решения, где неправильно переведенные фразы привели к отклонению доказательств или неправильному пониманию юридических вопросов.

Предлагается установить минимальные критерии точности машинных переводов в высоко рискованных областях, чтобы минимизировать риск ошибок, которые могут иметь серьезные последствия. Хотя трудно достичь абсолютной точности перевода из-за сложности языка и различий в оценке, важно разрабатывать стратегии для лучшего использования этих технологий и информированности о их потенциальных ограничениях.

В следующей главе я подробно рассматривал этические аспекты использования машинного перевода и влияние этой технологии на общество. Нейронные машинные переводы представляют собой значительный прогресс в преодолении языковых барьеров, однако с этим связаны и этические вопросы, такие как справедливое использование данных, защита личных данных и уважение авторских прав создателей контента.

Технологии машинного перевода, основанные на больших данных, отражают ценности и предположения тех, кто их разрабатывает и тренирует. Это может привести к риску искажения результатов и воспроизведению существующих неравенств. Поэтому я рассматривал возможности, как эти технологии могут выделять "хорошие" ценности и как они могут быть разработаны таким образом, чтобы минимизировать этические проблемы, такие как предвзятость и дискриминация.

Еще одной важной проблемой является владение данными. Хотя законное использование существующих данных для обучения систем является обычным, это вызывает вопросы, касающиеся прав создателей этих данных. Эти вопросы включают не только юридические, но прежде всего этические аспекты, такие как справедливая компенсация и прозрачность в их использовании.

В контексте машинного перевода крайне важно рассмотреть, какие последствия могут иметь технологические приложения для различных групп людей и как можно обеспечить, чтобы технологии использовались способом, который уважает права человека и способствует устойчивому развитию. Необходимо

рассмотреть, каким образом данные могут быть использованы или злоупотреблены, и какие меры можно принять для защиты тех, кто может быть отрицательно затронут технологиями.

Исследование и дискуссия об этике машинного перевода должны направляться на формулирование четких этических правил, которые обеспечат, что развитие и использование машинных переводов будут проводиться таким образом, который соответствует общепринятым моральным принципам и который способствует позитивному социальному воздействию технологий на общество.

В следующей части я сосредоточился на вопросах личных данных и конфиденциальности в контексте машинного перевода и онлайн-технологий. С 2018 года в ЕС действует регламент GDPR, который акцентирует внимание на защите личных данных и устанавливает строгие санкции за их нарушение. Этот регламент играет ключевую роль в повышении кибербезопасности и ограничивает использование серверов за пределами ЕС. Важно, чтобы вся вторичная переработка личных данных была осуществлена с согласия, предоставленного для первоначальных целей, что имеет существенное влияние на то, как организации управляют и делятся личными данными.

Нарушения в области защиты данных могут иметь разрушительное воздействие на доверие потребителей, рыночную стоимость и общественное восприятие компаний. Поэтому некоторые компании могут быть мотивированы скрывать такие нарушения, хотя большая прозрачность могла бы укрепить общественное доверие.

В то же время, в контексте возрастающего использования онлайн-платформ перевода, переводчики становятся более уязвимыми к возможному злоупотреблению их личными данными. Эти платформы могут снижать контроль переводчиков над их собственными данными и позволять контроль за их рабочей активностью, что увеличивает риск злоупотребления данными.

В области чат-ботов, которые становятся все более распространенными во многих секторах, GDPR подчеркивает безопасность и правильное обращение с личными данными. Чат-боты, имеющие доступ к личным данным, должны соблюдать правила GDPR, что включает обеспечение адекватных мер безопасности против несанкционированного доступа.

По этим причинам необходимо, чтобы организации, разрабатывающие и эксплуатирующие технологии, работающие с личными данными, такие как чат-боты, постоянно осознавали свои обязанности согласно GDPR и гарантировали, что весь сбор, хранение и использование личных данных происходят таким образом, который защищает конфиденциальность пользователей и обеспечивает безопасность их данных.

Другая очень актуальная проблема – возрастающие опасения относительно "машиностроительной предвзятости" в контексте машинного перевода. Эта предвзятость отражает существующие социальные неравенства, включая гендерные и расовые предубеждения, которые неосознанно проецируются в статистические модели. Примеры, когда технологии искусственного интеллекта проявляли предвзятость, включают инциденты, такие как неправильное распознавание лиц различных этнических групп или неправильная классификация на основе расовой принадлежности.

Особенно тревожным является предвзятость по отношению к полу в машинном переводе, что было задокументировано в недавнем исследовании, которое использовало список рабочих мест от Федерального бюро статистики труда США для создания предложений на 12 гендерно-нейтральных языках. Эти предложения были переведены на английский язык с помощью Google Translate, и анализ использованных местоимений показал значительную тенденцию к использованию мужских местоимений, особенно в областях, где преобладают мужчины или которые традиционно воспринимаются как мужские. Это открытие показывает, что Google Translate может несоразмерно фаворизировать мужские местоимения, тем самым воспроизводя гендерные стереотипы.

Google Translate, будучи одной из крупнейших общедоступных платформ для машинного перевода, использует данные от своих пользователей и из официальных источников, что усиливает опасения относительно сохранения гендерной нейтральности в переводах. Хотя были предприняты попытки уменьшить гендерное искажение, такие как снижение числа стереотипных аналогий, все еще существуют вызовы, связанные с устранением этих предубеждений из переводческих моделей.

Языки, которые грамматически различают пол, могут неосознанно укреплять гендерные стереотипы. В английском языке, где можно достичь гендерной нейтральности, переводческие инструменты должны поддерживать эту

нейтральность, чтобы не способствовать дальнейшему укреплению гендерных различий в обществе.

В целом необходимо, чтобы разработчики технологий машинного перевода и исследователи интенсивно работали над улучшением алгоритмов так, чтобы они отражали равенство и разнообразие без предвзятости.

Я также занимался постредктированием, которое приобретает всё большее значение в современном мире перевода. Благодаря повышению качества машинного перевода и развитию продвинутых нейронных сетей его использование значительно растёт, что снижает потребность в ручном переводе, но одновременно увеличивает спрос на постредктирование текстов, переведенных искусственным интеллектом. Поэтому важно и далее развивать и оптимизировать эту область.

При оценке качества машинного перевода и его последующего постредктирования важно найти правильный баланс между скоростью и качеством. Различает легкое и полное постредктирование, что показывает, что разные уровни корректировки могут соответствовать различным требованиям. Изучение этих процессов и их оптимизация имеет решающее значение для улучшения переводческих услуг и позволяет более эффективно использовать машинный перевод в сочетании с человеческой оценкой и корректировкой.

Сам оригинальный переводимый текст предоставляет комплексный взгляд на многие аспекты этой технологии. Текст включает широкий спектр тем, от основных определений, истории, через текущее использование, до этических и юридических дилемм, связанных с искусственным интеллектом. Публикация систематически исследует технологический прогресс и его влияние на социальные структуры и международные отношения, а также возможное использование искусственного интеллекта в военном секторе и его потенциальные опасности.

Текст также подробно рассматривает различия между слабым и сильным искусственным интеллектом и представляет возможности, как эти технологии могут изменить рынок труда, здравоохранение, домашнюю жизнь и многие другие области. Текст также предоставляет статистические данные о развитии и внедрении искусственного интеллекта в различных отраслях промышленности.

Мой перевод ограничивается выбранными главами, которые касаются развития, текущего использования, международных аспектов искусственного интеллекта, а также главами, занимающимися этическими вопросами и будущими перспективами. Целью этой работы было предоставить детальный взгляд на

текущее состояние и возможное будущее развитие искусственного интеллекта, его значение для будущего общества и вклад в технологический прогресс.

Далее я анализировал текст с формальной и содержательной точек зрения, где я анализировал его композицию, структуры, лингвистические аспекты, такие как синтаксис, морфологию и лексику.

Текст разделен на структурированные главы, которые систематически описывают различные аспекты искусственного интеллекта. Введение в тексте очерчивает основную информацию об искусственном интеллекте, следующие главы затем занимаются историей, развитием и текущим и будущим использованием искусственного интеллекта. Главы структурированы таким образом, чтобы постепенно переходить от общей информации к более детальным анализам конкретных тем.

В заключение я также рассмотрел способы перевода текста с русского на чешский с использованием различных переводческих трансформаций. Целью перевода было достичь соответствия с оригинальным текстом как с точки зрения содержания, так и стиля, с акцентом на сохранение сложной терминологии и структуры текста.

Среди наиболее часто используемых трансформаций была транслитерация, то есть переписывание графической формы слова из исходного языка, которую я использовал в основном для собственных имен. Кроме того, это была трансформация порядка слов, которая была необходима из-за различного порядка слов в русском и чешском языках. Эта трансформация позволила мне адаптировать структуру предложений таким образом, чтобы она соответствовала чешскому стилю речи.

Другой часто используемой техникой была универбализация, когда я заменял многословное название однословным, что служило для сжатия текста и упрощения выражения. Напротив, мультивербализация означала расширение однословного названия на многословное, что помогало лучше объяснить некоторые концепции и уточнить значение.

Генерализация и калькование были другими методами, которые я использовал для упрощения или точного перевода терминологии и концепций. Эти методы позволяют адаптировать текст таким образом, чтобы он был понятен чешскому читателю, при этом сохраняя верность оригинальному значению.

Использование этих трансформаций необходимо для достижения точного и функционального перевода, который уважает как лингвистические нюансы исходного языка, так и целевого языка





## **Anotace**

**Příjmení a jméno autora:** Alexandr Oros

**Název katedry a fakulty:** Katedra slavistiky, Filozofická fakulta Univerzity Palackého v Olomouci

**Název magisterské práce:** Komentovaný překlad odborného textu z technologické oblasti: Sociální a mezinárodní dopady využití technologie umělé inteligence

**Vedoucí práce:** Mgr. Martina Pálušová, Ph.D.

**Počet znaků:** 199 652

**Počet příloh:** 2

**Počet titulů použité literatury:** 28 + 3 internetové slovníky

**Klíčová slova:** odborný styl, terminologie, překlad, překladové transformace, technologie, umělá inteligence, komentář, etika, osobní údaje

### **Charakteristika:**

Magisterská práce „Komentovaný překlad odborného textu z technologické oblasti: Sociální a mezinárodní dopady využití technologie umělé inteligence“ je zaměřena na překlad ruské publikace „Mezinárodní a sociální dopady využití technologie umělé inteligence“ z ruského do českého jazyka a jazykovou analýzu terminologie dané oblasti. Součástí práce je glosář odborných slov a termínů z této oblasti. Teoretická část se skládá z kapitol, které propojují překládaný text s umělou inteligencí v oblasti překladu, konkrétně pak strojového překladu. Praktickou částí je pak samotný překlad publikace. V příloze se nachází glosář a originální text.

## Bibliografie

1. De Almeida, Giselle a Sharon O'Brien. Analysing post-editing performance: correlations with years of translation experience. In: *Proceedings of the 14th annual conference of the European Association for Machine Translation*, 2010.
2. Hasal, M. et al. Chatbots: Security, privacy, data protection, and social aspects. *Concurrency Computat Pract Exper*, 2021, vol. 33. Dostupné z: <https://onlinelibrary.wiley.com/doi/10.1002/cpe.6426>.
3. Hendl, Jan. Jazykové modely a umělá inteligence. *Medsoft*, 2023. Dostupné z: [https://www.creativeconnections.cz/medsoft/2023/Medsoft\\_sbornik\\_2023\\_Hendl.pdf](https://www.creativeconnections.cz/medsoft/2023/Medsoft_sbornik_2023_Hendl.pdf).
4. Hutchins, John. The history of machine translation in a nutshell. Retrieved December, 2005. Dostupné z: <https://aclanthology.org/www.mt-archive.info/10/Hutchins-2014.pdf>.
5. Irfan, Muhammad. Machine Translation. Bahria University, Islamabad, 2017 [cit. 4.5.2024]. Dostupné z: [https://www.researchgate.net/publication/320730405\\_Machine\\_Translation](https://www.researchgate.net/publication/320730405_Machine_Translation).
6. Kenny, Dorothy (ed.). Machine translation for everyone: Empowering users in the age of artificial intelligence (Translation and Multilingual Natural Language Processing 18). Berlin: Language Science Press, 2022. Dostupné z: <http://langsci-press.org/catalog/book/342>.
7. Khoong, Elaine C. et al. The Discharge Instructions Translation Study: A Randomized Controlled Trial to Assess the Efficacy of Google Translate in a Clinical Setting. *JAMA Network*, 2019 [cit. 15.4.2024]. Dostupné z: <https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2725080>.
8. Koehn, Phillip. Draft of chapter 13: Neural machine translation. In: *Statistical machine translation* [online], 2017 [cit. 6.5.2024]. Dostupné z: <https://arxiv.org/pdf/1709.07809.pdf>.
9. Liu, Kanglang a Muhammad Afzaal. Artificial Intelligence (AI) and Translation Teaching: A Critical Perspective on the Transformation of Education. *International Journal of Education Sciences*, 2021, 33(1-3), s. 64-73. [cit. 6.5.2024]. DOI: 10.31901/24566322.2021/33.1-3.1159. Dostupné z: <https://doi.org/10.31901/24566322.2021/33.1-3.1159>.

10. Lopez, Adam. Statistical machine translation. *ACM Computing Surveys*, 2008, 40(3), s. 8. Dostupné z: <https://dl.acm.org/doi/10.1145/1380584.1380586>.
11. Novelty Textile, INC. v. Windsor Fashions, INC. No. CV 12-05602 DDP (MANx), Central District of California, 2013. Dostupné z: [https://scholar.google.co.uk/scholar\\_case?case=378347259667692690&hl=cs&as\\_sdt=0](https://scholar.google.co.uk/scholar_case?case=378347259667692690&hl=cs&as_sdt=0).
12. O'Brien, Sharon. Machine Translation and Cognition. *The Handbook of Translation and Cognition*, 2017, s. 311-331. [cit. 2024-05-06]. DOI: 10.1002/9781119241485.ch17.
13. O'Brian, Sharon, Laura Winther Balling, Michael Carl, Michel Simard a Lucia Specia (eds.). *Post-editing of machine translation: processes and applications*. Newcastle upon Tyne: Cambridge Scholars Publishing, 2014. 322 s. ISBN 978-1-4438-5476-4. Dostupné z: [https://www.academia.edu/14415762/Post\\_editing\\_of\\_machine\\_translation\\_Processes\\_and\\_applications](https://www.academia.edu/14415762/Post_editing_of_machine_translation_Processes_and_applications).
14. Patil, Sumant a Patrick Davies. Use of Google Translate in medical communication: evaluation of accuracy. *BMJ Publishing Group*, 2014 [cit. 16.4.2024]. Dostupné z: <https://www.bmj.com/content/349/bmj.g7392>.
15. Pekařová, Dana. Lingvistická a translatologická analýza odborného textu z oblasti mlékárenského průmyslu. [online], 2019. Magisterská práce. Masarykova univerzita, Filozofická fakulta. Dostupné z: <https://is.muni.cz/th/fbasd/?cop=3436473;verze=2017;info=1>.
16. Prates, Marcelo O.R., Pedro H. Avelar a Luís C. Lamb. Assessing gender bias in machine translation: a case study with Google Translate. *Neural Computing and Applications*, 2020, vol. 32(15), s. 6363-6381. Dostupné z: <https://link.springer.com/article/10.1007/s00521-019-04144-6>.
17. Ragni, V., Alonso, E., Vieira, N. L. Translation Productivity: Tools, Tracking, Ethics [online], 2022. [cit. 14.4.2024]. Dostupné z: [https://www.researchgate.net/publication/361798340\\_Translation\\_Productivity\\_Tools\\_Tracking\\_Ethics](https://www.researchgate.net/publication/361798340_Translation_Productivity_Tools_Tracking_Ethics).
18. Randhawa, Gurdeeshpal, Ferreyra, Mariella, Ahmed, Rukhsana, Ezzat, Omar, Pottie, Kevin. Using machine translation in clinical practice, 2013. Dostupné z: <https://www.cfp.ca/content/cfp/59/4/382.full.pdf>.

19. Špačková, Stanislava. *Rusko-česká textová a jazyková ekvivalence propriálních lexik* [online]. Brno, 2016 [cit. 8.4.2024]. Dostupné z: <https://is.muni.cz/th/en8dp/> Disertační práce. Masarykova univerzita.
20. Texas Bar Journal. *Lost in Translation* [online], 2019. Dostupné z: <https://lscpagepro.mydigitalpublication.com/publication/?m=21412&i=562664&p=14&ver=html5> [cit. 14.4.2024].
21. Vieira, Lucas Nunes, Minako O'Hagan a Carol O'Sullivan. *Understanding the societal impacts of machine translation: a critical review of the literature on medical and legal use cases*. University of Bristol, 2020 [cit. 14.4.2024]. Dostupné z: <https://doi.org/10.1080/1369118X.2020.1776370>.
22. Vargová, Dominika a Juraj Varga. Creativity and copyright law in the process of translation. *Jazykovedný časopis (Journal of Linguistics)*, 2023, Vol. 74, No. 3, s. 761–782. Dostupné z: <https://intapi.sciendo.com/pdf/10.2478/jazcas-2024-0021>.
23. Vieira, N. L. *Translation Productivity: Tools, Tracking, Ethics* [online], 2022. Dostupné z: [https://www.researchgate.net/publication/361798340\\_Translation\\_Productivity\\_Tools\\_Tracking\\_Ethics](https://www.researchgate.net/publication/361798340_Translation_Productivity_Tools_Tracking_Ethics).
24. Vysloužilová, Eva a Milena Machalová. *Cvičebnice překladu pro rusisty, I. díl: Politika, ekonomika*. Olomouc: Univerzita Palackého v Olomouci, 2011. 235 s. ISBN 978-80-244-2854-3.
25. Vychodilová, Z. *Введение в теорию перевода для русистов*. 1. vyd. Olomouc: Univerzita Palackého v Olomouci, 2013. 83 s. Skripta. ISBN 978-80-244-3417-9.
26. Wang, Haifeng et al. *Progress in Machine Translation* [online]. Amsterdam: Elsevier, 2021 [cit. 20.4.2024]. ISSN 2095-8099. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S2095809921002745>.
27. Xie, W. et al. Detecting Symptom Errors in Neural Machine Translation of Patient Health Information on Depressive Disorders: Developing Interpretable Bayesian Machine Learning Classifiers. *Front. Psychiatry*, 2021. Dostupné z: <https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsy.2021.771562/full>.
28. Žváček, Dušan. *Kapitoly z teorie překladu, díl 1: Odborný překlad*. Olomouc: Vydavatelství Univerzity Palackého, 1995, 51 s. ISBN 807067489X.

## **Internetové slovníky a jiné příručky**

1. LINGEA SLOVNÍKY. Slovníky.lingea.cz [online]. Dostupné z: <http://slovníky.lingea.cz>.
2. PŘEKLADAČ GOOGLE. Translate.google.cz [online]. Dostupné z: <https://translate.google.cz/>.
3. YANDEX.TRANSLATE. Translate.yandex.com [online]. Dostupné z: <https://translate.yandex.com/>.
4. CZECHENCY.ORG. [online]. Dostupné z: <https://www.czechency.org/>

## **Пříloha č. 1 – originální text**

### **Введение**

Искусственный интеллект (ИИ), еще недавно существовавший лишь в научной фантастике, на сегодняшний день является одной из наиболее перспективных и быстро развивающихся технологий. Технологии ограниченного или «слабого» искусственного интеллекта уже активно используются в самых разных областях, от мобильных телефонов и бытовой электроники до военной продукции.

В то же время найдется мало сфер научного знания, настолько мифологизированных, как ИИ. Благодаря образам из популярной культуры искусственный интеллект часто предстает в СМИ едва ли не ящиком Пандоры, открыв который, человечество неизбежно окажется на грани апокалипсиса. Хотя подобный алармизм представляется беспочвенным, на повестке дня уже сегодня состоит разработка «сильного» искусственного интеллекта, способного к самостоятельному принятию осознанных управленческих решений.

Перспектива создания подобной технологии в значительной степени ставит под вопрос не только текущую систему глобального разделения труда, но и существующий мировой порядок и архитектуру международной безопасности. В условиях обострения противоречий и кризиса доверия между великими державами существует реальная опасность возникновения новой гонки вооружения в сфере технологий искусственного интеллекта. При этом, если контроль за оружием массового уничтожения – ядерным, химическим и биологическим – регулируется соответствующими международными договорами и конвенциями, то разработка военных технологий искусственного интеллекта на сегодняшний день остается в «серой зоне» международного права.

В то же время, на сегодняшний день очевидно, что искусственный интеллект будет все активнее использоваться в военной и гражданской сфере, в том числе в вопросах обеспечения стратегической стабильности. Соблазн получить новое более совершенное оружие и вырваться вперед в технологической гонке слишком велик, чтобы страны могли всерьез учитывать соображения общегуманитарного характера.

Одной из проблем, стоящих перед исследователями, является большая понятийная и терминологическая путаница, существующая в сфере ИИ. Данная рабочая тетрадь систематизирует существующий объем знаний о технологиях искусственного интеллекта и оценивает их перспективы их развития. В центре анализа – возможное влияние ИИ и смежных с ним технологий (машинное обучение, автономные аппараты) на международные отношения и различные сферы общественной жизни. В публикациях также рассматриваются этико-юридические аспекты применения технологий ИИ.

Рабочая тетрадь Российского совета по международным делам (РСМД) включает в себя аналитические материалы, подготовленные специалистами в сферах искусственного интеллекта, машинного обучения, автономных систем, а также юристов и социологов. Материал призван внести вклад в общественную дискуссию по вопросам искусственного интеллекта и возможных последствий применения данной технологии.

## **Искусственный интеллект: время слабых**

Шефтелович Д.Р.

Под понятием «искусственный интеллект» (ИИ), как правило, понимают компьютерную программу, т.е. алгоритм, способную на решение задач, которые в состоянии решить мозг взрослого человека. В Международном терминологическом словаре по искусственному интеллекту понятие «искусственный интеллект» определяется как область знаний, рассматривающая разработку технологий, позволяющих вычислительным системам действовать таким образом, которое напоминает разумное поведение, в том числе, поведение человека. Стоит обратить внимание на то, что это феноменологическое определение, оставляющее детализацию понятий «интеллект», «разумное поведение» на откуп философии; в условиях дефицита знаний о мозге и когнитивном аппарате биологических систем понятие ИИ не поддается более точной, математической формализации.

На текущий момент под исследованиями в области искусственного интеллекта сейчас принято понимать алгоритмическое решение проблем, требующих



когнитивного напряжения. Такими проблемами были (и отчасти остаются) игры: шахматы или Го, распознавание почерка, машинный перевод, творческая деятельность. В общественном сознании каждая из этих проблем первоначально казалась последней преградой к созданию «настоящего» ИИ, способного заменить человека во всех сферах его деятельности. В действительности оказалось, что, во-первых, научить компьютер играть в шахматы гораздо проще, чем научить его играть в футбол, и, во-вторых, даже зная, как научить компьютер играть в футбол, сложно перенести этот опыт на проблему машинного перевода. Поэтому после начального энтузиазма научное сообщество разделило понятие ИИ на категории сильного и слабого ИИ. Под слабым искусственным интеллектом понимается алгоритм, способный решать узкоспециализированные проблемы (вроде игры в шахматы), сильный искусственный интеллект же менее ограничен в спектре задач; в идеале сильный искусственный интеллект способен как минимум на то же, на что способен взрослый человек (например, делать логические выводы и планировать действия). Вместо понятия сильного ИИ в литературе также используется общий искусственный интеллект (Artificial General Intelligence, AGI).

Необходимо отметить, что классификация алгоритмических проблем на проблемы сильного и слабого ИИ не является исторической константой; так, на начало 2016 г. считалось, что для игры в Го нужен сильный ИИ. После впечатляющей победы алгоритма AlphaGo4 над действующим чемпионом мира в марте 2016 г. игра в Го стала проблемой слабого ИИ.

К концу 2010-х годов можно выделить следующие магистральные направления исследований в области алгоритмов искусственного интеллекта.

### ***Машинное обучение и распознавание образов***

Поиск закономерностей в данных, например, классификация объектов на фотографии по категориям «фон», «человек», «автомобиль», «здание», «растение».

### ***Планирование и логический вывод***

Доказательство логических утверждений и планирование действий для достижения определённой цели при знании логических закономерностей, позволяющих эту цель

достичь. Пример: синтез данных из сенсоров для оценки дорожной ситуации и адекватного управления автомобилем.

### ***Экспертные системы***

Систематизация данных вместе с логическими связями, отображение знаний, ответы на семантические запросы вроде «Какова доля цен на энергоносители в стоимости производства самолёта МС-21?».

### **Текущее состояние и перспективы**

Проблема создания сильного ИИ на данный момент не решена, и научное сообщество существенно охладело к теме ИИ после так называемой «зимы искусственного интеллекта», наступившей в середине 1980-х из-за завышенных ожиданий. Это привело, с одной стороны, к разочарованию потенциальных пользователей и к недостаточной производительности программных комплексов, с другой. Начиная с середины 2000-х годов удалось добиться значительных успехов в более специализированных задачах. В первую очередь это обусловлено постоянным развитием вычислительной техники: если в 2001 г. пределом возможностей был «флагманский» процессор Intel Pentium III, выполнявший 1,4 миллиарда арифметических операций в секунду, то, спустя десять лет, в 2011 году актуальный на тот момент процессор Intel Core i5 выполнял уже 120 миллиардов операций в секунду, т.е. почти в 100 раз больше. Во многом драйвером роста стала индустрия компьютерной графики и видеоигр: в погоне за всё более реалистичной графикой видеокарты из периферийного оборудования превратились в мощные вычислительные системы, способные не только обрабатывать графические данные, но и выполнять произвольные (пусть и не очень сложные) параллельные вычисления. Так, в том же 2011 году пиковая производительность видеокарты Radeon HD 6970 составляла 675 миллиардов операций в секунду.

Вторым важным фактором развития отрасли стала оцифровка и ручная классификация текстов, фотографий и звуковых записей, а также построение всеобъемлющих баз знаний в цифровом формате. Благодаря ставшим доступными большим объёмам качественно классифицированных данных стало возможно тренировать алгоритмы машинного обучения на больших выборках. Таким образом, точность классификации выросла, а машинный перевод превратился из грубого инструмента в систему широкой применимости. Так, известный большинству

пользователей переводчик компании Google работает за счёт огромной базы параллельных текстов, т.е., текстов идентичного значения на разных языках, дополняемой пользователями, что позволяет обучать систему и улучшать качество перевода прямо во время работы.

В сумме оба фактора позволили решать за приемлемое время всё более сложные задачи, сократив время обработки накопленных массивов данных. На начало 2018 г. искусственный интеллект добился нескольких ярких успехов, освещённых СМИ, проник в различные отрасли экономики. Однако было бы ошибкой считать ИИ «серебряной пулей», средством, способным решить все проблемы человечества.

## **Успехи ИИ**

### ***Экспертные системы и распознавание текстов: своя игра IBM Watson***

Пожалуй, наибольший общественный резонанс произвела экспертная система IBM Watson, объединившая в себе огромный массив знаний, т.е., данных с семантическими связями между ними, и способность обрабатывать запросы на английском языке к этой базе знаний. В 2011 году впечатляющим успехом IBM Watson была победа над действующими чемпионами в телевикторине «Jeopardy!». Этот успех позволил IBM успешно выйти на рынок экспертных систем, наглядно продемонстрировав способности Watson к обработке и структурированию информации.

### ***Нейросети: Победа в Го с Google DeepMind***

Во многом выиграла от повышения производительности вычислений технология искусственных нейронных сетей, известная ещё с середины 1940-х годов. Современные технологии позволяют обучать большие нейросети, а от размера сети прямо зависит сложность решаемых задач. Характерной чертой этих систем глубинного обучения является структура сети, позволяющая распознавать сначала локальные детали входных данных (вроде разницы в цвете между соседними пикселями), а по ходу обработки и их глобальные свойства (вроде линий и форм). Построив и натренировав нейросеть с подобной архитектурой, инженеры Google смогли сконструировать компьютерный алгоритм игры в Го, неожиданно выигравший у чемпиона мира. Схожих результатов другие группы исследователей добились, научив нейросети играть в компьютерные игры вроде StarCraft II.

## **Практическое применение ИИ**

На начало 2018 года машинное обучение, распознавание образов и автономное планирование вышли из исследовательских лабораторий в мир коммерческих приложений. Первыми пользователями новой технологии стали военные по обе стороны железного занавеса, начиная с 1950-х годов заинтересовавшиеся в решении задач планирования. Практически одновременно с ними задачи планирования стали решать экономисты. Приведём несколько примеров применения алгоритмов искусственного интеллекта.

- IBM предлагает услуги упомянутого выше Watson различным отраслям: медикам для диагностики симптомов пациента и рекомендации терапии, юристам для классификации конкретных ситуаций в соответствии с правовыми нормами, железнодорожникам для оценки усталости составов и путей.
- В области медицины распознавание образов позволяет опознать и классифицировать органы для планирования хирургических процедур.
- Онлайн-магазины пользуются механизмами машинного обучения для более точной рекомендации товаров регулярным клиентам.
- Автономные роботизированные музейные гиды способны проводить экскурсии и отвечать на вопросы посетителей по теме экспозиции.
- В военной сфере уже сейчас можно увидеть зачатки автономного принятия решений: комплексы ближней противовоздушной и противоракетной обороны исключают человека из цепочки принятия решений из-за низкой скорости реакции человеческих операторов, а современные противокорабельные ракеты способны распределять цели между ракетами в залпе в зависимости от их важности.

## **Перспективы применения ИИ**

Ожидается, что в 2020-х годах ИИ сможет решать ещё больше задач. Ниже перечислим некоторые из них и ориентировочный прогресс научно-исследовательских и опытно-конструкторских работ.

### ***Автономные транспортные средства***

Наиболее сложной задачей для автономных транспортных средств считается участие в дорожном движении в связи с ограниченной свободой манёвра и множеством возможных ситуаций, возникающих на дороге. В 2018 г. самым

высокоразвитым решением на рынке считается автопилот Audi A8, способный автономно двигаться в условиях пробок на автострадах. В марте 2017 г. представители BMW обещали к 2021 году представить автомобиль, способный без вмешательства водителя доставлять пассажиров в место назначения.

### ***Военное применение***

Самым перспективным направлением для военных считается возможность автоматического распознавания и сопровождения целей для роботизированных платформ, и, как следствие, автономное принятие решения об их поражении. На конец 2017 года в странах «первого эшелона» проводится ряд научно-исследовательских работ по этим темам, в первую очередь в отношении наземных, надводных и подводных комплексов.

### ***Роботизированные хирургические процедуры***

Ожидается, что хирургия, в том числе и микрохирургия, требующая точного вмешательства в тело пациента, сможет быть автоматизирована в ближайшем будущем. В 2016 г. был представлен в качестве демонстратора технологий робот STAR, способный проводить операции на мягких тканях. При достижении успехов в распознавании образов можно ожидать роботизации всех хирургических процедур удаления тканей, что позволит сделать хирургию более доступной и снизить нагрузку на лечащий персонал.

### **Нерешённые проблемы**

Даже для кажущихся всемогущими компьютерных алгоритмов существуют нерешаемые и плохо решаемые проблемы. В частности, по очевидным причинам сложно решить задачи, трудно поддающиеся формализации, например, «написать роман» или «выбрать самую красивую фотографию». Но даже при наличии математически точной формулировки нет гарантии успешного решения. Причиной этому может быть сложность математического моделирования проблем более низкого уровня (например, моделирование движения при обучении робота играть в футбол), сложность самой проблемы (например, для логического вывода и доказательства математических утверждений нет алгоритмов существенно лучше полного перебора возможных логических цепочек), огромное количество параметров и неточность наблюдаемого мира (например, при игре в футбол), и недостаточная производительность вычислительных систем по сравнению с человеческим мозгом. В конце концов, взаимодействие  $1 \times 10^{11}$  человеческих

нейронов не поддается алгоритмической симуляции; на начало 2018 года наибольшим успехом в этой области была симуляция  $1,7 \times 10^9$  нейронов с 2400-кратным

замедлением в 2013 году; при этом нет уверенности, что симуляция даже нужного количества клеток мозга позволит воссоздать мозговую деятельность в компьютерной модели.

Отдельно стоит упомянуть недостатки машинного обучения. Как правило, машинное обучение предполагает наличие уже классифицированных тренировочных данных, в которых компьютерный алгоритм находит закономерности. Из-за недостаточных данных в работе алгоритма возможны ситуации, когда данные на входе не принадлежат ни к одному из тренированных классов; распознавание нового феномена на входе и автоматическое создание нового класса объектов считается как минимум нетривиальной

задачей; её можно легко усложнить, добавив в условие активное обучение во время работы классификатора и временное изменение распознаваемых классов. Вторым существенным недостатком машинного обучения считается сверхчувствительность: так, распознавание лиц можно «перехитрить», надев кажущиеся безобидными очки. В отдельных случаях можно добиться неправильной классификации фотографии путём невидимых человеку изменений: так, после кажущейся несущественной манипуляции

панда на фотографии может классифицироваться как обезьяна.

В то время, как компьютеры успешно решают «сложные» задачи вроде символических и численных вычислений и обыгрывают гроссмейстеров в шахматы, в списке нерешённых проблем много задач, отличающихся относительно простой формулировкой: классификация «неизвестных» образов без тренировки на заранее классифицированных образцах (например, распознавание яблок на фотографии, если известны только классы «вишня» и «груша»), моторика, рассуждения от «здравого смысла». В литературе этот феномен известен как «парадокс Моравека». Во многом это отображает человеческое восприятие: создававшиеся миллионами лет эволюцией способности, присутствующие у любого взрослого человека, кажутся само собой разумеющимися, в то время как математические проблемы

вроде нахождения кратчайшего пути на карте кажутся неестественными и их решение малоочевидным.

### ***Опасности ИИ***

В апокалиптических сценариях научно-фантастических произведений искусственный интеллект, как правило, в какой-то момент решает уничтожить человечество, а учёные не в состоянии этому помешать либо не осознают эту опасность. На практике в научном сообществе вопрос опасности сверх-интеллекта обсуждается достаточно давно. Основной опасностью считается неправильное целеполагание «сильному» ИИ, располагающему существенными вычислительными и материальными ресурсами, и не принимающему во внимание интересы людей. Прочитанная выше книга Бострома предлагает возможные решения этой проблемы.

### ***Реальность сильного ИИ***

Завышенные либо заниженные ожидания на успехи научно-технического прогресса влекут за собой прогнозы, в которых сильный ИИ либо станет доступным уже завтра (в крайнем случае – в следующем году), либо принципиально недостижим. Фактически создание сильного искусственного интеллекта сложно спрогнозировать, так как оно зависит от успешного решения нескольких инженерных и математических задач неизвестной сложности. Исторический пример решения теоремы Ферма, появившегося спустя 350 лет после её формулировки, не позволяет с уверенностью прогнозировать сроки решения задач такого масштаба.

### **Международные тенденции**

Расширение границ применимости ИИ привело к повышенному интересу военных и околвоенных кругов к возможностям автономных систем. Помимо НИОКР по упомянутым выше темам, в международном сообществе разгорелась дискуссия о возможном ограничении или даже запрете робототехнических комплексов. Наибольшую известность получила кампания «Stop Killer Robots», активно требующая полного запрета на разработку автономных боевых систем по этическим причинам. Однако в этой связи стоит упомянуть не только боевых роботов, но и системы классификации, влияющие на решение о применении силы исключительно по метаданным, даже не обращая внимания на содержание писем подозреваемых.

Общественное внимание к автономным боевым системам привело к переговорам в рамках Конвенции ООН о запрещении или ограничении применения конкретных видов обычного оружия, которые могут считаться наносящими чрезмерные повреждения или имеющими неизбирательное действие, под действие которой подпадают противопехотные мины и ослепляющие лазеры. Однако на начало 2018 года дипломатический процесс не привёл к каким-либо взаимным обязательствам. Во многом это обусловлено

сложностью в определении «автономной боевой системы», наличием комплексов (в первую очередь ПВО и ПРО), удовлетворяющих возможному определению, и нежеланием отказываться от перспективной технологии. Тем не менее, было бы излишним упрощением ограничивать аргументы за контроль над автономными вооружениями морально-этическими аспектами. Во многом в пользу ограничений говорят аспекты стратегической стабильности: во-первых, применение автономных систем может привести к неконтролируемой эскалации военных конфликтов с непредсказуемыми последствиями; во-вторых, сложно контролировать экспорт используемых технологий: если экспорт крылатой ракеты очевиден и её дальность ограничена физическими параметрами, то контроль такого уровня над программным кодом невозможен: область применения алгоритмов не ограничена одной лишь военной сферой, а ограничение исследований в области с настолько широким спектром применения, вплоть до запрета публикаций, неизбежно натолкнётся на сопротивление научного сообщества, во многом живущего за счёт международной кооперации и крупных конференций.

## **Искусственный интеллект – благо или угроза для человечества?**

Колонин А.Г.

### **Основные проблемы развития и потенциальные угрозы искусственного интеллекта**

Несмотря на множество современных достижений в нейробиологии, на сегодняшний день пока никто точно не знает, как устроен естественный интеллект. Соответственно, точно так же никто не знает, как именно создать искусственный



интеллект. Существует ряд известных проблем, требующих решения для его создания и разные мнения по поводу приоритетности достижения тех или иных решений. Например, руководитель международных проектов по созданию искусственного интеллекта с открытым кодом OpenCog и SingularityNET Бен Герцель считает, что все необходимые технологии для создания общего ИИ в принципе уже разработаны, необходимо только соединить их некоторым правильным образом для получения такой синергии, результатом которой станет возникновение общего ИИ. Другие эксперты настроены более скептически, полагая, что необходимо принципиальное решение многих проблем, которые будут перечислены ниже. Также сильно варьируются экспертные оценки срока возникновения сильного ИИ – от десятка до нескольких десятков лет.

Возникновение систем даже просто автономного или адаптивного, а тем более общего или сильного ИИ связывается с несколькими угрозами разного масштаба, которые актуальны уже сегодня. Среди них можно отметить:

- создание так называемых «автономных систем смертоносных вооружений» – Lethal Autonomous Weapons Systems (LAWS), например, дронов для заказных убийств;
- новый виток гонки вооружений, при котором государства будут совершенствовать уровни интеллекта автономных средств поражения и уничтожения;
- интеллектуальная система, не обязательно боевая, но и промышленная или бытовая, способна не только к целенаправленному действию, но и к сознательному целеполаганию, при том, что автономная постановка целей системы может привести к постановке целей, противоречащих целям человека;
- исключение из материального производства подавляющей части населения за счет еще большей автоматизации, что может приводить к еще большему социальному расслоению, снижению эффективности «социальных лифтов» и увеличению массы «лишних людей» с соответствующими социальными последствиями.

Сложность контроля за системами искусственного интеллекта на сегодняшний день обусловлена, в частности, «закрытостью» существующих прикладных решений на основе «глубоких нейронных сетей», которые не позволяют не только верифицировать правильность принятия решений перед их исполнением, но даже по факту проводить анализ решения, которое было принято машиной. Решению этой

проблемы сейчас посвящено как новое направление «объяснимый искусственный интеллект» (Explainable Artificial Intelligence, EAI), так и новый интерес к интеграции ассоциативного (нейро-сетевого) и символьного (основанного на логике) подходов к проблеме.

В августе 2018 г. в Праге на площадке Чешского Технического Университета одновременно прошли конференции, посвященные искусственному интеллекту человеческого уровня, общему искусственному интеллекту, биологически-вдохновленным когнитивным архитектурам, а также нейро-символьным технологиям. На конференциях были представлены доклады ведущих специалистов компаний и учреждений, лидирующих в сфере исследовательских разработок в области искусственного интеллекта (ИИ): Microsoft, Facebook, DARPA, MIT, Good AI. В этих докладах были обозначены как текущее состояние разработок в области ИИ, так и стоящие перед обществом нерешенные проблемы, а также угрозы, возникающие в ходе дальнейшего развития этой технологии. В этом обзоре я постараюсь кратко обозначить основные проблемы и угрозы, а также возможные пути противостояния этим угрозам.

Однако, прежде всего, необходимо уточнить значение некоторых терминов, которые обычно употребляются совместно с ИИ в различных контекстах: слабый или специализированный ИИ, автономный ИИ (Autonomous AI), адаптивный ИИ (Adaptive AI), общий ИИ (Artificial General Intelligence, AGI), сильный ИИ (Strong AI), ИИ человеческого уровня (Human-Level AI), ИИ сверхчеловеческого уровня (Super-human AI).

**Слабый или специализированный ИИ** представлен всеми без исключения существующими решениями и предполагает способность автоматизации решения одной конкретной задачи, будь то игра в Go или распознавание лиц на видеокамерах. При этом отсутствует возможность самостоятельного обучения другим задачам без перепрограммирования человеком.

**Автономный ИИ** предполагает возможность системы функционировать долгое время без участия оператора. Например, он позволяет дрону, оборудованному

солнечными батареями, совершить многодневное путешествие с Елисейских полей на Красную площадь или в обратном направлении, самостоятельно выбирая как маршрут, так и места для промежуточных посадок для подзарядки аккумуляторов, избегая при этом всевозможные препятствия.

**Адаптивный ИИ** предполагает способность системы адаптироваться к новым условиям, приобретая знания, не закладываемые при создании. Например, позволить системе поддержания диалогов на русском языке самостоятельно осваивать новые языки и применять их знание в разговоре, попадая в новую языковую среду или на основе изучения учебных материалов для этих языков.

**Общий ИИ** предполагает настолько высокую адаптивность, что обладающая им система может быть использована в самых различных видах деятельности при соответствующем обучении. Обучение может быть как самостоятельным, так и направленным (с помощью инструктора). В этом же смысле в противопоставление слабому или специализированному ИИ также часто употребляется сильный ИИ.

**ИИ человеческого уровня** предполагает уровень адаптивности сравнимый с человеческим, то есть система способна осваивать те же самые навыки, что и человек в сопоставимые сроки обучения.

**ИИ сверхчеловеческого уровня** предполагает еще более высокую адаптивность и скорость обучения. Таким образом, система может обучиться тем знаниям и способностям, которые человеку в принципе не под силу.

### **Принципиальные проблемы создания сильного ИИ**

Возникновение сильного искусственного интеллекта вполне закономерно в рамках общего эволюционного процесса, как закономерно и возникновение молекул из атомов, клеток из молекул, организмов из клеток, выделение специализированных клеток в центральную нервную систему, возникновение социальных структур, развитие речи, письменности и в конечном итоге – информационных технологий. Закономерность нарастающей сложности информационных структур и способов организации в процессе эволюции хорошо показана Валентином Турчиным. Если не произойдет гибель человеческой цивилизации, то такая эволюция будет неизбежной, и в самой долгосрочной перспективе это будет спасением

человечества, постольку только не биологические формы существования информации смогут пережить неизбежную со временем гибель Солнечной системы и смогу сохранить во Вселенной информационный код нашей цивилизации.

При этом важно осознавать, что для того, чтобы построить сильный искусственный интеллект, не обязательно понимать, как устроен естественный, так как необязательно понимать, как летает птица, чтобы сделать ракету. Очевидно, это будет сделано рано или поздно тем или иным способом или, возможно, несколькими способами.

В качестве принципиальных проблем, решение которых еще предстоит для создания общего или сильного ИИ большинство экспертов выделяют следующие:

- **Быстрое обучение** (few-shot learning) – необходимость построения систем, обучающихся на небольшом объеме материала в отличие от существующих систем глубокого обучения, требующих большие объемы специально подготовленного обучающего материала.
- **Сильная генерализация** (strong generalisation) – создание технологий распознавания ситуаций, в которых распознаваемые объекты встречаются в условиях, отличных от тех, в которых они встречались в использованном для обучения материале.
- **Генеративные или генерирующие модели обучения** (generative models) – разработка технологий обучения, когда объектом запоминания являются не признаки объекта распознавания, а принципы его формирования, что может позволить отражать более глубокие сущности распознаваемых объектов и осуществлять более быстрое обучение и более сильную генерализацию.
- **Структурированное обучение и предсказание** (structured prediction and learning) – развитие технологий обучения на основе представления объектов обучения в виде многослойных иерархических структур, где более низкоуровневые элементы определяют более высокоуровневые, что может оказаться альтернативным решением проблем быстрого обучения и сильной генерализации.

- Решение проблемы **катастрофического забывания** (catastrophic forgetting) – присущего большинству существующих систем, которые, будучи изначально обучены на одном классе объектов и затем до-обучены распознаванию на новом классе объектов, теряют способность распознавать объекты первого класса.
  
- Достижение **возможности инкрементального обучения** (incremental learning), предполагающего способность системы накапливать знания и совершенствовать свои возможности постепенно, не теряя при этом полученные ранее знания, но приобретая новые знания применительно к системам диалогового общения на естественном языке. Идеальным является прохождение «младенческого теста Тьюринга» (Baby Turing Test), в случае которого система должна продемонстрировать возможность постепенного освоения языка от уровня младенца до уровня взрослого человека.
  
- Решение проблемы **сознания** (consciousness) – предполагает формирование проверенной рабочей модели сознательного поведения, обеспечивающего эффективное прогнозирование и целенаправленное поведение за счет формирования «внутренней картины мира» в рамках которой возможен поиск оптимальных стратегий поведения по достижению поставленных целей без фактического взаимодействия с реальным миром, что существенно повышает безопасность, проверки гипотез, а также повышает скорость и энергетическую эффективность этой проверки, тем самым создавая возможность самообучения живой или искусственной системы в «виртуальном мире» собственного сознания. С прикладной точки зрения проблема сознания имеет две стороны. С одной – создание систем ИИ, обладающих сознанием, позволит резко повысить их эффективность. С другой – появление сознания у таких систем вызывает как дополнительные риски, так и вопросы этического плана, поскольку такие системы в какой-то момент смогут быть по уровню самосознания приравнены к самому человеку с вытекающими из этого последствиями в правовом поле.

## **Потенциальные угрозы ИИ**

Возникновение систем даже просто автономного или адаптивного, а тем более общего или сильного ИИ связывается с несколькими угрозами разного масштаба, актуальными уже сегодня.

Во-первых, угрозу для человека может представлять интеллект не обязательно сильный, общий, человеческого или сверх-человеческого уровня, так как достаточно иметь автономную систему, оперирующую большими объемами информации с большими скоростями. На её основе могут быть созданы так называемые «автономные системы смертоносных вооружений» – Lethal Autonomous Weapons Systems (LAWS), простейший пример которых – дроны для заказных убийств, печатаемые на 3D-принтерах как в массовых масштабах, так и небольшими партиями в кустарных условиях.

Во-вторых, угрозу для государства может представлять ситуация, когда другое государство (потенциальный противник) получает вооружения с более адаптивным, автономным и общим искусственным интеллектом с повышенной скоростью реакции и предсказательной способностью.

В-третьих, угрозу для всего мира представляет вытекающая из предыдущей угрозы ситуация, когда государства вступают в новый виток гонки вооружений, совершенствуя уровни интеллекта автономных средств поражения и уничтожения – как это было предсказано Станиславом Лемом несколько десятков лет назад.

В-четвертых, угрозу для любой стороны представляет любая, не обязательно боевая, но и промышленная или бытовая интеллектуальная система с определенной степенью автономности и адаптивности, способная не только к целенаправленному действию, но и к сознательному целеполаганию, при том что автономная постановка целей системы может привести к постановке целей, противоречащих целям человека и людей, а возможностей достижения этих целей у системы будет намного больше, в силу её более высокого быстродействия, большего объема обрабатываемой информации и большей предсказательной способности. К сожалению, масштабы именно этой угрозы сообществом не вполне изучены и осознаны.

В-пятых, угрозу для общества представляет переход к новому уровню развития производственных отношений в капиталистическом (либо тоталитарном) обществе, когда более малочисленная часть населения получает возможность контролировать материальное производство, исключая из него подавляющую часть населения за счет еще большей автоматизации, что может приводить к еще большему социальному расслоению, снижению эффективности «социальных лифтов» и увеличению массы «лишних людей» с соответствующими социальными последствиями.

Наконец, угрозу для человечества в целом может представлять автономизация глобальных вычислительных систем обработки данных, распространения информации и принятия решений на основе глобальных сетей, поскольку скорость распространения информации в таких системах и масштаб воздействия может приводить к непредсказуемым с позиций имеющегося опыта и существующих моделей управления социальным явлениям. Например, внедряемая система социального кредита в современном Китае является уникальным экспериментом цивилизационного масштаба с непонятными на сегодняшний день последствиями.

Сложность контроля за системами искусственного интеллекта на сегодняшний день обусловлена, в частности, «закрытостью» существующих прикладных решений на основе «глубоких нейронных сетей», которые не позволяют не только верифицировать правильность принятия решений перед их исполнением, но даже по факту проводить анализ решения, которое было принято машиной. Решению этой проблемы сейчас посвящено как новое направление «объяснимый искусственный интеллект» (Explainable Artificial Intelligence, EAI), так и новый интерес к интеграции ассоциативного (нейро-сетевого) и символического (основанного на логике) подходов к проблеме.

## **Дивный новый мир без работы**

Маркоткин Н.М.

Что приходит нам в голову, когда мы думаем о грядущем повсеместном внедрении роботов и технологий искусственного интеллекта (ИИ)? Уходящие за горизонт очереди за пособием по безработице? Дроны «Скайнета», бороздящие небо над выжженными трущобами? Или же наоборот, тотальная праздность и коммунизм, обеспеченный трудом механических рабов? Реальность, скорее всего, будет не такой яркой, однако это не повод не задуматься о социальных последствиях происходящих на наших глазах технологических изменений.

### **Революция на марше**

Мир стремительно вступает в стадию IV промышленной революции, среди ключевых компонентов которой роботизация, развитие био- и нанотехнологий, 3D печати, Интернета вещей, генетики и технологий искусственного интеллекта. Грядущие технологические изменения будут иметь прямые последствия для целого ряда существующих профессий и полностью трансформируют рынок труда как минимум в развитых странах.

Высокий темп происходящих перемен (достаточно сказать, что 10 из наиболее востребованных в 2010 г. специальностей не существовали в 2004 г.) затрудняет прогнозирование их воздействия на общество. В этой связи оценки экспертов и международных организаций варьируются от оптимистических, до алармистских. Однако даже если отбросить радикальные сценарии, можно с уверенностью говорить о том, что в обозримом будущем нас ждет фундаментальная перестройка глобальной экономики, сравнимая с той, что произошла в XVIII–XIX вв., во времена I промышленной революции.

Согласно докладу Всемирного экономического форума (ВЭФ) «Будущее профессий», 65% учащихся сегодня в начальных классах будут иметь абсолютно новые, не существующие ныне профессии. К аналогичным выводам пришла компания McKinsey, в докладе которой подчеркивается, что при текущем уровне развития технологий 60% профессий содержат 30% функционала, который может



быть автоматизирован. М. Осборн и К. Фрей из Оксфордского университета дают еще более пессимистичный прогноз. Согласно результатам их исследования, 47% рабочих мест в США рискуют быть автоматизированными в течение 20 лет.

### **Кого заменят роботы**

Какие же профессии находятся в зоне риска? В первую очередь речь идет, конечно же, о неквалифицированном рутинном труде. В исследовании М. Осборна и К. Фрея в число тех, кто наиболее уязвим, попали клерки, наборщики данных, библиотекари, операторы станков, сантехники, специалисты по продажам, наладчики оборудования и другие.

Согласно оценкам ВЭФ, с 2015 по 2020 гг. сокращение рабочих мест в наибольшей степени коснется офисных профессий (4,91%) и производственного сектора (1,63%). В пределах 1% также должна сократиться занятость в таких сферах, как дизайн, индустрия развлечений, строительство и продажи. В свою очередь, наиболее значительный прирост рабочих мест предсказуемо ожидается в сфере компьютерных технологий (3,21%), архитектурных и инженерных специальностях (2,71%) и менеджменте (чуть менее 1%).

Под угрозой автоматизации в среднесрочной перспективе также предсказуемо окажутся профессии, связанные с транспортом. Развитие технологии самоуправляемых автомобилей способно радикальным образом изменить рынок как пассажирских, так и в особенности грузовых перевозок. Только в США в индустрии дальних грузовых перевозок заняты 8,7 млн человек.

Если же учитывать весь бизнес, завязанный на дальнобойщиках (мотели, придорожные кафе и т. д.), то это число может возрасти до 15 млн, то есть порядка 10% от трудоспособного населения страны. Едва ли не более значимыми могут стать сокращения и в секторе пассажирских перевозок, общественного транспорта. Велика вероятность того, что самоуправляемые технологии в ближайшем будущем будут также внедрены в морских грузовых перевозках. По мере развития технологий искусственного интеллекта тяжелые времена ожидают также юристов, учителей,

шахтеров, менеджеров среднего звена, журналистов и представителей целого ряда других профессий.

Можно говорить о том, что основная занятость постепенно переместится из сферы услуг в другие секторы экономики, многие из которых еще только предстоит создать. Впрочем, подобная перспектива не уникальна – она скорее является подтверждением революционного характера происходящих перемен. До I индустриальной революции более 70% населения было занято в сельском хозяйстве, сегодня же этот показатель в развитых странах колеблется в районе нескольких процентов. Доля занятых в промышленности росла вплоть до середины XX в., однако в результате Цифровой революции на сегодняшний день она снизилась до 24% в ЕС и 19% в США (в России этот показатель составляет 27%). При этом, хотя рабочих становится меньше, объем производства продолжает стабильно расти. Теперь же, похоже, пришла очередь автоматизации услуг.

### **Золотой век инженеров и психиатров?**

Среди профессий, которые с наименьшей вероятностью пострадают в краткосрочной перспективе, чаще всего называют те, которые связаны с интеллектуальным трудом или подразумевают непосредственный персональный контакт с клиентом. К примеру, в исследовании Оксфордского университета к наиболее защищенным от автоматизации профессиям отнесены различные виды врачебных и психологических специальностей, а также тренеры, социальные работники, программисты, инженеры, представители высшего управленческого звена и творческих профессий.

Иными словами, лучшим образом подготовлены к новым условиям те, чья работа требует творческого подхода и не сводится к выполнению определенных отработанных комбинаций. В этом плане, говоря, к примеру, об инженерах, важно уточнить, что в «безопасности» находятся в первую очередь инженеры-проектировщики, в то время как инженеры-эксплуатационники, наоборот, находятся в зоне риска.

Автоматизацию креативных профессий сдерживают три ключевых фактора. Для успешного выполнения их задач искусственный интеллект должен обладать

восприятием и возможностью оперировать материальными объектами (осознанием), творческим интеллектом и социальным интеллектом. Текущий уровень технологического развития действительно не позволяет решить эти задачи. Однако по мере развития технологий «сильного» ИИ спектр доступных ему работ будет неуклонно расширяться. Он позволит расширить пределы автоматизации, которые уже были достигнуты при существующих технологиях, и даст компьютерам возможность принимать управленческие решения и даже, возможно, заниматься творческой деятельностью. Поэтому нельзя исключать, что в среднесрочной или долгосрочной перспективе машины смогут успешно заменить не только инженеров и менеджеров, но и писателей и художников. Тем более что прецеденты успешного написания ИИ художественных текстов уже существуют.

Таким образом, вполне возможно, что большей части трудоспособного населения предстоит в обозримом будущем снова сесть за парту. Однако проблема заключается в том, что никто толком не знает, чему именно им придется учиться. По некоторым оценкам, вплоть до 85% профессий, которые будут востребованы в 2030 г., на сегодняшний день еще не существуют. Текущие же системы образования даже в развитых странах еще не адаптировались к новой реальности.

## **Этические и правовые вопросы искусственного интеллекта**

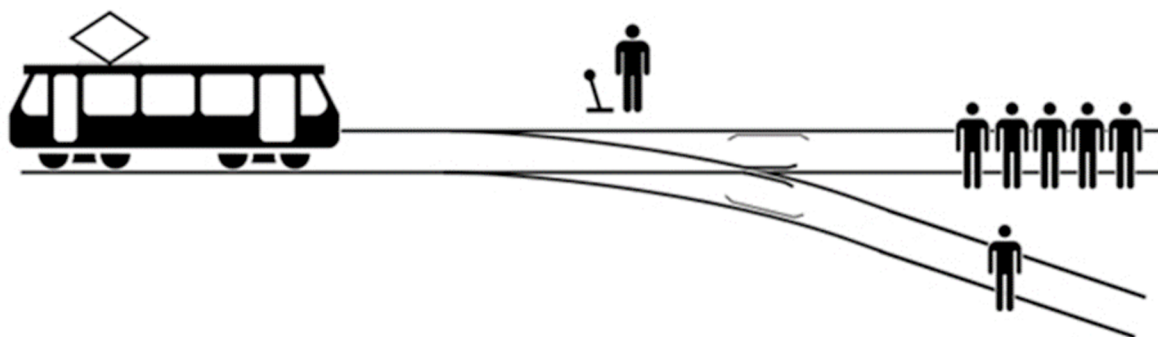
Карлюк М.В.

Этика и право неразрывно связаны в современном обществе, и многие правовые решения вытекают из восприятия тех или иных этических проблем. Искусственный интеллект добавляет новое измерение в данные вопросы. Системы, использующие технологии искусственного интеллекта, становятся всё более автономными в смысле сложности задач, которые они могут выполнять, их потенциального влияния на мир и уменьшающейся способности человека понимать, предсказывать и контролировать их функционирование. Большинство людей недооценивает реальную автономность таких систем. Они могут учиться на собственном опыте и осуществлять действия, которые не были задуманы их создателями. Этим

обусловлен ряд этических и правовых затруднений, которые будут затронуты в данной статье.

### **Этика и искусственный интеллект**

Существует довольно известный эксперимент – проблема вагонетки. Он поднимает ряд важных этических вопросов, имеющих непосредственное отношение к искусственному интеллекту. Представьте, что неуправляемая вагонетка несётся по рельсам, а на ее пути к рельсам привязаны пять человек. Вы стоите возле рычага, с помощью которого можно переключить стрелку, и вагонетка повернёт и поедет по другому пути, где к рельсам привязан один человек. Переключите ли Вы стрелку?



Однозначного ответа на данный вопрос нет. Более того, вариаций ситуации, в которой приходится принимать подобное решение, очень много. К тому же разные социальные группы дают разные ответы. Так, например, буддистские монахи преимущественно готовы пожертвовать жизнью одного человека, чтобы спасти пятерых даже в более сложном варианте проблемы вагонетки.

Что касается искусственного интеллекта, такая ситуация может возникнуть, например, на дороге, по которой движется беспилотный транспорт, в случае, если авария непреодолима. Возникает вопрос, чья жизнь должна быть в приоритете – пассажиров, пешеходов или ни тех, ни других? В Массачусетском технологическом университете был даже создан специальный сайт, посвящённый этой проблеме, где пользователь может попробовать на себе различные сценарии и выбрать, как стоило бы поступить в той или иной ситуации.

Возникают также вопросы, что позволено делать с правовой точки зрения? На основе чего делать выбор? И кто в итоге будет виноват? Данная проблема уже получила отклики от компаний и регуляторов. Так, представители компании

«Мерседес» прямо сказали, что их машины будут отдавать приоритет пассажирам, на что в Министерстве транспорта Германии был сразу же дан ответ, предвосхищающий будущее регулирование, – что делать такой выбор на основе ряда критериев будет неправомерно и в любом случае производитель будет нести ответственность.

Другие страны могут выбрать иной путь. Возьмём, к примеру, китайскую Систему социального рейтинга, в соответствии с которой каждому гражданину присваивается рейтинг в зависимости от законопослушности, полезной общественной деятельности и др. Обладатели низких рейтингов будут подвергаться санкциям. Что помешает руководству Китая заставить производителей беспилотного транспорта предусмотреть правило, что в случае непреодолимой аварии жертвовать необходимо человеком с наименьшим рейтингом? За счёт, например, технологий распознавания лиц и доступа к соответствующим базам данных, вполне возможно определить и сравнить рейтинги потенциальных жертв.

### **Основные проблемы, стоящие перед правом**

Но правовые проблемы ещё глубже, особенно в случае роботов. Система, которая учится на информации, полученной из внешнего мира, может действовать таким образом, который создатели не могли предсказать, а предсказуемость критична для современных правовых подходов. Более того, такие системы могут действовать независимо от создателей или операторов, что усложняет задачу по определению субъекта ответственности. Данные характеристики ставят проблемы предсказуемости и способности действовать самостоятельно, но не быть юридически ответственным.

Варианты регулирования многочисленны, в том числе на основе уже существующих норм. Так, технологии, использующие искусственный интеллект, можно регулировать как объекты авторских прав либо как имущество. Однако сложности возникают, если учесть, в частности, способность к автономному действию помимо воли создателя, собственника или владельца. В связи с этим можно применить нормы, регулирующие особый вид имущества – животных, поскольку последние также способны к автономным

действиям. В российском праве к животным применяются общие правила об имуществе (ст. 137 ГК РФ), поэтому ответственность будет наступать по ст. 1064 ГК РФ: вред, причинённый личности или имуществу гражданина, подлежит возмещению в полном объёме лицом, причинившим вред, которым является собственник автономного агента.

Предложения о применении правового режима животных высказываются, однако они имеют некоторые ограничения. Во-первых, применение законодательства по аналогии недопустимо в рамках уголовного права. Во-вторых, данные нормы созданы в первую очередь для домашних животных, которые не должны причинять вред при обычных обстоятельствах. Для более развитых систем существуют предложения провести аналогию с дикими животными, поскольку для последних существуют более жёсткие правила. Однако и здесь возникает вопрос, как сделать такое разделение в связи с особенностями искусственного интеллекта, указанными выше. Более того, жёсткие правила могут привести к замедленному внедрению технологий искусственного интеллекта в связи с большими и непредсказуемыми рисками ответственности для создателей.

Распространённым предложением является применение к таким системам норм о юридических лицах. Поскольку юридическое лицо является искусственно сконструированным субъектом права, то роботов также можно наделить аналогичным статусом. Право может быть достаточно гибким и наделять практически любыми правами. Оно также может и ограничить в правах. Например, исторически рабы не обладали практически никакими правами и вообще являлись имуществом. Можно также наблюдать и противоположную ситуацию, когда объекты без явных признаков способности к действию наделяются правами. Даже сегодня существуют примеры необычных объектов, признающихся юридическими лицами, – как в развитых, так и в развивающихся странах. В Новой Зеландии в 2017 г. был принят закон, который признал статус юридического лица за рекой Уонгануи. В законе указывается, что данная река является юридическим лицом и обладает всеми правами, полномочиями и обязанностями юридического лица. Таким образом, данный закон трансформировал реку из собственности в юридическое лицо, что расширило границы понимания относительно того, что может являться собственностью, а что нет. В Индии Верховный

Суд в 2000 г. признал юридическим лицом основной священный текст сикхов – Гуру Грантх Сахиб.

Но даже если не рассматривать экстремальные варианты, а привести в пример обычные компании, правовые системы предусматривают гражданско-правовую, а в некоторых случаях и уголовно-правовую ответственность юридических лиц. Не определяя, что компании (или государства) обладают свободой воли или интенцией, или что они могут действовать преднамеренно или осознанно, считается возможным признавать их юридически ответственными за определённые действия. Таким же образом, не обязательно приписывать интенцию или свободу воли роботам, чтобы признавать их ответственными за то, что они делают.

Однако аналогия с юридическими лицами проблематична, поскольку, по сути, концепция юридического лица нужна для быстрого и эффективного осуществления правосудия, но действия юридических лиц всегда восходят к действиям индивида или группы людей, даже если невозможно их точно определить. Другими словами, правовая ответственность компаний и похожих образований связана с действиями, выполняемыми их представителями или работниками. Более того, в случае уголовной ответственности, когда она признаётся за юридическими лицами, она возможна только при условии, что физическое лицо, совершившее действие от имени юридического лица, определено. Действия же систем на основе искусственного интеллекта не обязательно будут прямо восходить к действиям человека.

Наконец, к таким системам можно применять правовые нормы об источниках повышенной опасности. В соответствии с п. 1 ст. 1079 ГК РФ юридические лица и граждане, деятельность которых связана с повышенной опасностью для окружающих (использование транспортных средств, механизмов и др.), обязаны возместить вред, причинённый источником повышенной опасности, если не докажут, что вред возник вследствие непреодолимой силы или умысла потерпевшего. Проблема заключается в разграничении, какие системы могут быть отнесены к источникам повышенной опасности. Этот вопрос похож на вышеуказанную проблему разграничения домашних и диких животных.

## **Национальное и международное регулирование**

Многие страны активно создают правовые условия для развития технологий, использующих искусственный интеллект. Так, в Южной Корее ещё с 2008 г. существует Закон «О содействии развитию и распространению умных роботов». Закон направлен на улучшение качества жизни и развитие экономики путём разработки и продвижения стратегии устойчивого развития индустрии умных роботов, а правительство каждые пять лет разрабатывает основной план обеспечения эффективного достижения данной цели.

Здесь хотелось бы остановиться на двух недавних примерах – Франции, которая заявила об амбициях стать европейским и мировым лидером в сфере искусственного интеллекта; и Европейском союзе, в котором предложены продвинутые нормы регулирования умных роботов.

### ***Франция***

В конце марта 2018 г. президент Франции Э. Макрон презентовал национальную стратегию в сфере искусственного интеллекта. Франция планирует инвестировать 1,5 млрд евро в течение пяти лет для поддержки исследований и инноваций в данной сфере. Стратегия основана на рекомендациях, сделанных в отчёте, подготовленном под руководством французского математика и депутата национального собрания Франции Седрика Виллани. Причём сделан выбор направить стратегию на конкретные четыре сектора: здравоохранение, транспорт, окружающую среду и оборону, и безопасность. Это сделано для того, чтобы использовать потенциал сравнительных преимуществ и сфер компетенций с фокусом на секторах, в которых компании смогут играть ключевую роль на глобальном уровне, а также из-за их важности для общественного интереса и др.

В целом даны семь ключевых предложений, одно из которых представляет особый интерес для целей статьи – сделать искусственный интеллект более открытым. Действительно, алгоритмы закрыты и в большинстве случаев являются коммерческой тайной. Однако алгоритмы могут быть предвзятыми, например, в процессе самообучения впитать стереотипы, существующие в обществе либо



переданные разработчиками, и на их основе принимать решения. Таким образом уже принимаются судебные решения. В США подсудимый был приговорён к продолжительному сроку заключения на основе информации, полученной от алгоритма, оценивающего

возможность повторного преступления. Подсудимый безуспешно оспаривал использование алгоритма для принятия такого решения, поскольку не были предоставлены критерии оценки, являющиеся коммерческой тайной. Французская стратегия предлагает развить прозрачность алгоритмов и возможностей по их проверке, а также определить этическую ответственность работающих в сфере искусственного интеллекта, создать консультативные комитеты по этике и т. д.

### ***Европейский союз***

В ЕС первым шагом в направлении регулирования вопросов искусственного интеллекта стала Резолюция Европейского Парламента 2017 г. «Нормы гражданского права о робототехнике». Ещё в 2015 г. в Европейском Парламенте была создана рабочая группа по правовым вопросам, связанным с развитием робототехники и искусственного интеллекта в ЕС. Резолюция не является обязательным документом, но дает ряд рекомендаций Европейской Комиссии для возможных действий в этом направлении, причём не только касательно норм гражданского права, но и этических аспектов робототехники.

Резолюция определяет «умного робота» как робота, получающего автономию через использование сенсоров и/или взаимосвязь с окружающей средой; при этом робот имеет по меньшей мере минимальную физическую поддержку, адаптирует своё поведение и действия в соответствии с условиями среды, при этом у него отсутствует жизнь с биологической точки зрения. Предлагается создать систему регистрации продвинутых роботов, которая управлялась бы Агентством ЕС по робототехнике и искусственному интеллекту. Данное агентство также предоставляло бы техническую, этическую и регулятивную экспертизу по робототехнике. Что касается ответственности, предлагаются два варианта: либо объективная ответственность (не требующая вины), либо подход риск-менеджмента (ответственность лица, которое могло минимизировать риски). Ответственность должна быть пропорциональной реальному уровню указаний, которые отдаются роботу и уровню его автономности. Правила ответственности могут быть

дополнены обязательным страхованием для пользователей роботов, и компенсационным фондом для выплаты компенсации в случае отсутствия страхового полиса, покрывающего риск.

Для решения этических вопросов резолюция предлагает два кодекса поведения: Кодекс этики для разработчиков робототехники и Кодекс комитетов по этике научных исследований. Первый, в частности, предлагает четыре этических принципа: 1) «делай благо» (роботы должны действовать в интересах людей); 2) «не навреди» (роботы не должны причинять вред человеку); 3) автономия (взаимодействие человека с роботами должно быть добровольным); 4) справедливость (выгоды, получаемые от деятельности роботов, должны быть распределены справедливо).

Таким образом, примеры, приведённые в статье, показывают, среди прочего, как общественные ценности влияют на отношение к искусственному интеллекту и его юридическому оформлению. Поэтому наше отношение к автономным системам, будь то роботы или что-то другое, наше переосмысление их роли в обществе и положения среди нас, может иметь трансформационный эффект. Правосубъектность определяет то, что важно для общества, и позволяет решить, является ли «что-то» ценным и надлежащим объектом для обладания правами и обязанностями.

В связи с особенностями искусственного интеллекта звучат предложения о непосредственной ответственности определённых систем. В соответствии с такими взглядами, не существует фундаментальных причин, почему автономные системы не должны нести ответственность за свои действия. Однако открытым остаётся вопрос о необходимости либо желательности введения такой ответственности (во всяком случае на данном этапе). Это в том числе связано с этическими проблемами, указанными в статье. Возможно, сделать программистов или пользователей автономных систем ответственными за действия таких систем будет более эффективным. Однако это же может замедлить инновации. Поэтому необходимо продолжить поиски оптимального баланса.

Для этого важно ответить на многие вопросы. Например, какие цели мы преследуем, развивая технологии искусственного интеллекта? Чьи интересы должны преобладать? Насколько эффективно это будет? Ответ на эти вопросы поможет предотвратить ситуации подобно той, которая случилась в России XVII в., когда животное (а именно, козёл) было отправлено в ссылку в Сибирь за совершённые им действия.

## **Пříloha č. 2 – Terminologický glosář**

### **A**

- **Алгоритм** – algoritmus
- **Алармизм** – alarmismus
- **Аналитические модели** – analytické modely
- **Адаптивные системы** – adaptivní systémy
- **Автономные системы** – autonomní systémy
- **Автономные аппараты** – autonomní zařízení
- **Ассоциативный подход** – asociální přístup

### **B**

- **Взаимодействие человеческих нейронов** – interakce lidských neuronů
- **Виртуальный мир** – virtuální svět

### **Г**

- **Генеративные модели** – generativní modely
- **Генерализация** – generalizace
- **Глубокие нейронные сети** – hluboké neuronové sítě
- **Гибридные модели** – hybridní modely

### **Д**

- **Данные** – data
  - **Систематизация данных** – systematizace dat
  - **Доступ к данным** – přístup k datům

- **Диагностика** – diagnostika
- **Динамическое планирование** – dynamické plánování

## **И**

- **ИИ** – UI, umělá inteligence
  - **Сильный ИИ** – silná UI
  - **Слабый ИИ** – slabá UI
  - **Общий ИИ** – obecná UI
  - **ИИ человеческого уровня** – UI lidské úrovně
  - **ИИ сверхчеловеческого уровня** – UI nadlidské úrovně
- **Информационные технологии** – informační technologie
- **Исследования** – výzkumy

## **К**

- **Контроль за ИИ** – kontrola UI
- **Когнитивный анализ** – kognitivní analýza
- **Компьютерная архитектура** – počítačová architektura

## **Л**

- **Логические выводы** – logické závěry

## **М**

- **Машинное обучение** – strojové učení
- **Машинный перевод** – strojový překlad
- **Модель** – model
- **Морфологическая структура** – morfologická struktura

## **Н**

- **Научно-технический прогресс** – vědecko-technický pokrok
- **Нейронные сети** – neuronové sítě
- **Нейросетевые технологии** – technologie neuronových sítí

## **О**

- **Обработка данных** – zpracování dat

- **Оцифровка** – digitalizace

## **П**

- **Программные комплексы** – softwarové systémy
- **Прогнозирование** – prognóza

## **Р**

- **Распознавание** – rozpoznávání
- **Роботизированные системы** – robotizované systémy

## **С**

- **Системы** – systémy
- **Семантические связи** – sémantické vazby

## **Т**

- **Технологии** – technologie
- **Тест Тьюринга** – Turingův test

## **Ф**

- **Формализация** – formalizace
  - **Математическая формализация** – matematická formalizace