UNIVERZITA
KOMENSKÉHO
V BRATISLAVE

Peter Šantavý

UMELÁ INTEL!GENCIA

DOBRÝ SLUHA A ZLÝ PÁN?

Ľudia, stroje, spoločnosť...
Etika, hodnoty, budúcnosť...

Druhé, rozšírené vydanie

UNIVERZITA KOMENSKÉHO V BRATISLAVE

RÍMSKOKATOLÍCKA CYRILOMETODSKÁ

BOHOSLOVECKÁ FAKULTA

Peter Šantavý

Umelá inteligencia – dobrý sluha a zlý pán?

Ľudia, stroje, spoločnosť...

Etika, hodnoty, budúcnosť...

Druhé, rozšírené vydanie

Bratislava, 2025

Vydavateľ

Vydala Rímskokatolícka cyrilometodská bohoslovecká fakulta

Univerzity Komenského v Bratislave v roku 2025.

Druhé, rozšírené vydanie

Rozsah: ~~309~~ 404 strán (pdf); ~~18~~ 25 AH

Brožovaná verzia (paperback):	ISBN	978-80-88696-95-7
	EAN	9788088696957
Elektronická verzia (pdf):	ISBN	978-80-88696-96-4
	EAN	9788088696964

Autor

ThLic. Ing. Peter Šantavý, PhD.

Katolícky kňaz, správca linuxových systémov a počítačových sietí, expert v oblasti kybernetickej bezpečnosti, nechtiac technologický manažér a príležitostný učiteľ IKT (informačných a komunikačných technológií).

Má viac ako 30 rokov skúseností s IKT, pohybuje sa v oblasti kybernetickej bezpečnosti, komplexných počítačových systémov a sieťových technológií.

Fenoménu umelej inteligencie sa venuje s osobitným zreteľom na oblasť kybernetickej bezpečnosti, etiky, informačnej spoločnosti a AGI.

Catholic priest, GNU/Linux and computer network administrator, cybersecurity expert, (unintentionally) technology manager and occasional ICT (Information and Communication Technology) teacher.

He has more than 30 years of experience in ICT, working in cyber security, complex computer systems and network technologies.

He works on the phenomenon of artificial intelligence, with a particular focus on cybersecurity, ethics, the information society and AGI.

Recenzenti

doc. ThDr. Ing. Vladimír Thurzo, PhD., RKCMBF UK Bratislava

Mgr. Marek Zeman, PhD. CRISC, FIIT STU Bratislava

Knihu napísal sám autor.

Žiadna časť tohto diela nebola vygenerovaná strojovým algoritmom:-)

Abstrakt

Publikácia predstavuje určité nóvum, snažiac sa o inovatívny prístup v nazeraní na problematiku umelej inteligencie v kontexte kresťanského svetonázoru a etických dôsledkov limitov a rizík, s ktorými sa táto atraktívna oblasť moderného technologického sveta v súčasnosti potýka: interdisciplinárne uchopenie fenoménu umelej inteligencie, dôsledná analýza limitov i rizík a z nej prameniace etické závery, návrh základnej štruktúry všeobecných i špecifických etických princípov a zásad, vyjadrenie diskutabilných faktorov i niektorých perspektív všeobecnej umelej inteligencie smerujúcej k superinteligencii a zdôvodnenie jasného postoja k jej porovnávaniu s ľudskou bytosťou – to všetko tvorí vklad do prebiehajúceho celosvetového etického diskurzu, ktorý tým viac naberá na dôležitosti, čím viac sa technológie umelej inteligencie rozvíjajú a ich sofistikované implementácie sa do reálneho života masívne zavádzajú, ovplyvňujúc tak životy miliónov ľudí.

Publikácia vychádza z autorovej dizertačnej práce *Etické výzvy a morálne aspekty súčasných systémov umelej inteligencie**, pričom už v prvom vydaní** bol pôvodný text upravený a podstatne rozšírený o ďalší výskum v oblasti všeobecnej umelej inteligencie.

Čitateľovi sa dostáva do rúk druhé, rozšírené vydanie, v ktorom sa autor snaží nielen doplniť niektoré parciálne poznatky, technické parametre a novinky, ale predovšetkým reflektovať vývoj, ktorý nastal po predstavení najnovších generatívnych systémov umelej inteligencie, pričom ich neustále narastajúci spoločenský dopad a dôsledky si pred rokom ešte takmer nikto nevedel predstaviť.

———————

* ŠANTAVÝ, P. *Etické výzvy a morálne aspekty súčasných systémov umelej inteligencie* [dizertačná práca]. [on-line]. Bratislava: RKCMBF UK, 2022. Dostupné na internete: <https://peter.santavy.cloud/data/uploads/docs/eticke_vyzvy_a_moralne_aspekty_sucasnych_systemov_umelej_inteligencie.pdf>

** ŠANTAVÝ, P. *Umelá inteligencia – dobrý sluha a zlý pán?* [on-line]. Bratislava: RKCMBF UK, 2023. ISBN 978-80-88696-91-9. Dostupné na internete: <https://peter.santavy.cloud/ai-good-servant-and-bad-master>

1. Základy umelej inteligencie v skratke

Umelá inteligencia je ako Colombova žena, nikto ju nevidel a všetci o nej hovoria.[53]

I keď umelú inteligenciu pokladáme za niečo nové a bytostne sa viažuce na súčasnú modernú – primárne informačnú – spoločnosť, ide o fenomén, ktorý oslovoval mysliteľov a vizionárov už niekoľko storočí dozadu, možno od počiatkov Prvej priemyselnej revolúcie[54], keďže jedným z podstatných prvkov priemyselnej revolúcie bolo nahradenie ľudskej práce strojmi. Činnosť, ktorú dovtedy manuálne vykonávali ľudia, odrazu vykonávali – a efektívnejšie vykonávali – stroje. S konkrétnou paradigmatickou zmenou vtedajšej spoločnosti sa tak začali otvárať nové horizonty aj v pohľade na zariadenia, ktoré môže človek vytvoriť.

Zaujímavým predstaviteľom tohto snaženia môže byť Kempelenov[55] Turek, mechanický stroj zostrojený v roku 1771, ktorý bol vydávaný za šachový automat. Stroj, ktorý bol prezentovaný ako automat, dokázal zohrať solídnu šachovú partiu proti ľudskému protihráčovi. V skutočnosti však išlo o mechanické zariadenie ukrývajúce skutočného šachistu, pričom podvod vyšiel na povrch až v roku 1857, t.j. niekoľko rokov po požiari, pri ktorom stroj zhorel (r. 1854).[56]

Nerozoberajúc mechanické prevedenie Turka[57] je možné si uvedomiť **ideu vytvorenia**

53 doc. Ing. Ľuboš MAGDOLEN, CSc., SjF STU.

54 V rámci Prvej priemyselnej revolúcie sa ako dôsledok pokroku vo výrobných nástrojoch (osobitne v textilných manufaktúrach) a následne vynálezu parného stroja začína industrializácia a masívne využívanie strojov, ktoré nahrádzali ručnú prácu.
Priemyselná revolúcia [on-line]. [cit. 20. augusta 2020].
Dostupné na internete: <https://sk.wikipedia.org/wiki/Priemyseln%C3%A1_revol%C3%BAcia>

55 Barón Wolfgang Kempelen, pochádzajúci z nemecky hovoriacej rodiny, bol uhorský vysoký štátny úradník, polytechnik a vynálezca, ktorý sa narodil 23. januára 1734 v Bratislave.
Wolfgang Kempelen [on-line]. [cit. 20. augusta 2020].
Dostupné na internete: <https://sk.wikipedia.org/wiki/Wolfgang_Kempelen>

56 *Turek (stroj)* [on-line]. [cit. 20. augusta 2020].
Dostupné na internete: <https://sk.wikipedia.org/wiki/Turek_(stroj)>

57 V súčasnosti Amazon ponúka službu *Mechanical Turk* ako „trhovisko pre prácu, ktorá vyžaduje ľudskú inteligenciu“. Mechanical Turk od Amazonu spája záujemcov s úlohami ťažko realizovateľnými počítačmi so zamestnancami, ktorí za poplatok tieto úlohy vykonajú. Takto sa napr. pripravujú veľké množiny dát (sorting and labeling in datasets) pre učenie a testovanie systémov umelej inteligencie.

stroja, ktorý hrá šach a dokáže sa v tejto hre, ktorá bola dlhé roky považovaná za prejav inteligencie, postaviť človekovi.

Tieto tendencie boli následne umocnené s príchodom Druhej a osobitne Tretej priemyselnej revolúcie, v rámci ktorej nové *mysliace stroje* (t.j. počítačové systémy) boli stále viac schopné vykonávať koncepčné, riadiace i administratívne funkcie a koordinovať tok výroby, od ťažby surovín po predaj a distribúciu konečných výrobkov i služieb.[58] Technologické zariadenia a informačné systémy sa tak v podstatnej miere začali podieľať na fungovaní spoločnosti a jej premene na spoločnosť informačnú[59].

Príchod Tretej priemyselnej revolúcie (známej tiež ako digitálnej), ktorý nastal v polovici 20. storočia, znamenal aj úsvit umelej inteligencie a jej systematického bádania.

Digitálna revolúcia, ktorá bola počiatkom informačného veku a premeny spoločnosti na spoločnosť založenú na vedomostiach, sa v kontexte permanentnej technickej revolúcie[60] podieľa na formovaní v súčasnosti nastupujúcej Štvrtej priemyselnej revolúcie (známej ako *Industry 4.0*). Je charakterizovaná zlúčením technológií, ktoré stierajú hranice

---

MITCHELL, *Artificial Intelligence*, s. 85.

58 Druhá priemyselná revolúcia spadá do obdobia druhej polovice 19. storočia až po 1. svetovú vojnu. V tomto čase sa masívnym spôsobom rozrastá priemyselná výroba a prichádza k elektrifikácii spoločnosti.
Tretia priemyselná revolúcia, ktorá sa datuje niekedy od polovice 20. storočia a je tiež známa ako digitálna, je obdobím nástupu digitálnych technológií, počítačov a moderných foriem spracovania informácií i komunikácie. Digitálna revolúcia bola začiatkom informačného veku a premeny spoločnosti na spoločnosť založenú na vedomostiach.
HUMBER, M. *Technology and Workforce: Comparison between the Information Revolution and the Industrial Revolution* [on-line]. Berkeley: University of California, 2007, s. 2-3. [cit. 20. augusta 2020]. Dostupné na internete: <http://infoscience.epfl.ch/record/146804/files/InformationSchool.pdf>

59 ŠANTAVÝ, *Analýza virtuálneho sveta v kontexte náboženskej formácie, pôsobenia a života Cirkvi* [on-line], s. 19-20. [cit. 21. augusta 2020].
Dostupné na internete: <https://peter.santavy.cloud/data/uploads/docs/analyza_virtualneho_sveta_v_kontexte_nabozenskej_formacie_posobenia_a_zivota_cirkvi.pdf>

60 ŠANTAVÝ, *Analýza virtuálneho sveta v kontexte náboženskej formácie, pôsobenia a života Cirkvi* [on-line], s. 23-24. [cit. 20. augusta 2020].
Dostupné na internete: <https://peter.santavy.cloud/data/uploads/docs/analyza_virtualneho_sveta_v_kontexte_nabozenskej_formacie_posobenia_a_zivota_cirkvi.pdf>

medzi fyzickými, digitálnymi a biologickými sférami s dôrazom na celkovú transformáciu spoločnosti na spoločnosť znalostnú a informačnú, pričom dôležitú úlohu zohráva nástup systémov a prvkov umelej inteligencie do takmer všetkých oblastí fungovania spoločnosti.[61]

1.1. Úsvit umelej inteligencie

Ako už bolo spomenuté, úsvit umelej inteligencie nastával ruka v ruke s príchodom digitálnej revolúcie a s rozvojom elektronických výpočtových systémov.

Vtedajší vizionári a vynálezcovia v oblasti informačných technológií prekračovali základný i aplikovaný výskum v oblasti matematiky a informatiky pohľadom skôr filozofickým či futurologickým so základnou otázkou, kam až môže vývoj počítačových systémov zájsť.

Keďže informačné a komunikačné technológie (IKT) sa čoraz viac etablovali ako *mysliace stroje* (vysvetlenie a významové zaradenie je spomenuté v predchádzajúcej kapitole), k úvahám o umelej inteligencii bol už len malý krok, ktorý viacerí s nadšením a entuziazmom neváhali vykonať.

Medzi dôležité artefakty tohto obdobia bezpochyby patrí tzv. **Turingov test**, ktorého cieľom bolo zistiť, či je nejaký stroj inteligentný, resp. dokáže myslieť. Turingov test je založený na jednoduchej premise: ak dokáže človek aspoň päť minút konverzovať s nejakým respondentom bez toho, že by zistil, že ide o stroj (nie človeka), tak tento stroj (systém, počítač,...) úspešne prejde testom.[62] Inak povedané, dokáže stroj napodobniť ľudské myšlienky, dokáže myslieť?

---

61 *Prečo Industry 4.0* [on-line]. [cit. 20. augusta 2020].
Dostupné na internete: <http://industry4.sk/o-industry-4-0/co-je-industry-4-0/>

62 V Turingovom teste osoba v úlohe pýtajúceho sa dáva otázky dvom respondentom, s ktorými nie je v priamom kontakte, pričom jedným respondentom je človek a druhým stroj. Komunikácia je výlučne textová, trvá päť minút a rozsah konverzačných tém nie je obmedzený.
Ak pýtajúci sa spolu s hodnotiacou komisiou (hodnotiť odpovede môže viac osôb) nedokáže podľa odpovedí respondentov rozlíšiť, ktorý z nich je človek a ktorý stroj, potom tento stroj možno považovať za inteligentný. Prelomovou hranicou sa považuje 30% úspešnosť, teda minimálne traja z desiatich hodnotiteľov musia byť presvedčení, že komunikácia prebieha s človekom a nie strojom, prípadne nedokážu rozlíšiť človeka od stroja.
Por. *Turingov test* [on-line]. [cit. 29. januára 2021].
Dostupné na internete: <https://sk.wikipedia.org/wiki/Turingov_test>

Test sa prvý krát podarilo zvládnuť v roku 2014 počítačovému programu Eugene, ktorý simuloval konverzáciu trinásťročného chlapca.[63] A po ňom nasledovali ďalšie úspešné programy, čo len dokazuje, že Turingov test nie je dostatočným kritériom pre umelý systém, ktorý rozmýšľa ako človek. **Základným problémom – a jeho podstata sa týka prakticky takmer všetkých systémov umelej inteligencie dneška – je skutočnosť, že i keď nejaký systém dokáže odpovedať tak, akoby konverzácii rozumel, neznamená to, že je inteligentný a že konverzácii aj skutočne rozumie**.[64] [65] [66]

Každopádne **Alan Turing[67], priekopník a otec modernej počítačovej vedy, ktorý svoj**

63 *First Turing Test success marks milestone in computing history* [on-line]. [cit. 29. januára 2021]. Dostupné na internete: <https://phys.org/news/2014-06-turing-success-milestone-history.html>

64 Realitu tohoto problému vystihuje aj argument čínskej izby, ktorý bol predložený filozofom Johnom Searlom v roku 1980. Ide o myšlienkový experiment názorne vyjadrujúci, že **schopnosť zmysluplne odpovedať na položené otázky nie je dostatočná na preukázanie schopnosti otázkam porozumieť a myslieť**.
SEARLE, R. J. *Minds, Brains, and Programs* In: *The Behavioral and Brain Sciences.* [on-line]. Cambridge University Press, 1980, zv. 3. [cit. 30. januára 2021].
Dostupné na internete: <https://web.archive.org/web/20071210043312/http://members.aol.com/NeoNoetics/MindsBrainsPrograms.html>

65 Nielen program Eugene, ale aj oveľa sofistikovanejšie systémy AI, napr. Watson od IBM, vytvárajú chyby, ktoré sú však často iné, než robia v podobných situáciách ľudia. Sú také „neľudské“, ľudsky nepochopiteľné. Je však zaujímavé, že jedným z faktorov, ktorý dokázal hodnotiteľov Eugene presvedčiť, bola schopnosť robiť chyby, ktoré sú podobné ľudským chybám.

66 Jedným zo spôsobov, ktorý sa pri systémoch spracovania jazyka, inteligentných asistentoch a pod. v súčasnosti používa na meranie schopnosti porozumieť komunikácii, je tzv. SQuAD (The Stanford Question Answering Dataset), ktorý bol v roku 2016 predstavený výskumníkmi Stanfordskej univerzity.
RAJPURKAR, P. et al. *SQuAD: 100 000+ Questions for Machine Comprehension of Text,* In: *Proceeding of the 2016 Conference on Empirical Methods in Natural Language Processing.* 2016, 2382-92.
Často využívaným spôsobom na testovanie inteligencie je i tzv. Winograd Schema Challenge. Patrí do skupiny testov, v rámci ktorých sa systému AI predkladajú rôzne tvrdenia (formou otázok s výberom odpovedí), pričom systém preukazuje schopnosť chápania správnym výberom odpovedí na jednotlivé otázky.
Commonsense Reasoning ~ Winograd Schema Challenge. [on-line]. [cit. 18. júl 2024].
Dostupné na internete: <https://commonsensereasoning.org/winograd.html>

67 Alan Turing (1912-1954) bol britský matematik, logik a kryptograf, ktorý súbežne (nezávisle od seba) s matematikom Johnom von Neumannom vytvoril princípy dnešných elektronických počítačov. Turing sa už počas II. svetovej vojny preslávil skonštruovaním zariadenia, ktoré umožnilo prekonať nemecké šifrovacie zariadenie Enigma. Povojnovú prácu na konštrukcii digitálnych počítačov rozšíril o štúdium

test predstavil v roku 1950, ním akoby položil základy umelej inteligencie a odštartoval niečo, čo sa už nedalo zastaviť.

Za skutočný zrod umelej inteligencie sa však považuje **Dartmouthský seminár** – dvojmesačný výskumný projekt zaoberajúci sa umelou inteligenciou, ktorý sa uskutočnil v lete 1956 v Dartmouth college v USA.[68] V podstate išlo o pracovný seminár, ktorý organizoval mladý matematik John McCarthy v spolupráci s Marvinom Minskym (bývalým kolegom zo štúdií, ktorý s ním zdieľal fascináciu inteligentnými počítačmi a neskôr sa stal zakladateľom laboratória umelej inteligencie na MIT), Claudem Shannonom (kolegom z Bell Labs/IBM a vynálezcom teórie informácií) a Nathanielom Rochesterom (pionierom v oblasti elektroniky a architektom viacerých počítačov IBM).

Prvé použitie, či **vynájdenie termínu *umelá inteligencia*** sa pripisuje práve Johnovi McCarthymu, ktorý sa tak chcel vyhraniť voči príbuznej, čerstvo sa rozvíjajúcej oblasti kybernetiky.[69] [70] Je pritom zaujímavé, že použitie prívlastku umelá sa vtedy nepáčilo prakticky nikomu na Dartmouthskom seminári, keďže mali na mysli pravú, skutočnú

---

neurónových sietí, uplatnenie matematických metód v biológii a skúmanie vízie mysliacich strojov. *Alan Turing* [on-line]. [cit. 30. januára 2021]. Dostupné na internete: <https://en.wikipedia.org/wiki/Alan_Turing#Pattern_formation_and_mathematical_biology> *The Turing Digital Archive* [on-line]. [cit. 30. januára 2021]. Dostupné na internete: <http://www.turingarchive.org/>

68 *Dartmouth workshop* [on-line]. [cit. 1. februára 2021]. Dostupné na internete: <https://en.wikipedia.org/wiki/Dartmouth_workshop>

69 Kybernetika je vo všeobecnosti veda o skúmaní, riadení, regulácii a komunikácii v dynamických systémoch v technike, biologickej sfére a spoločnosti. Predmetom kybernetiky je spracovanie informácií pre potreby riadenia a jej metódami sú systémový prístup a modelovanie pri riešení problémov. *Kybernetika* [on-line]. [cit. 3. februára 2021]. Dostupné na internete: <https://sk.wikipedia.org/wiki/Kybernetika>

70 Na margo vyhranenia sa voči kybernetike však treba povedať, že umelá inteligencia a kybernetika vo svojom rozvoji majú styčné plochy a spoločné oblasti, napr. robotika, automatizácia, počítačové videnie a pod. Autor tohto textu, absolvujúc štúdium technickej kybernetiky na Elektrotechnickej fakulte SVŠT (aktuálne pôsobiacej pod názvom Fakulta elektrotechniky a informatiky STU), v rámci už vtedajšieho kybernetického odboru navštevoval aj základy neurónových sietí a vo svojej diplomovej práci riešil spracovanie obrazu metódami matematickej morfológie, t.j. metódami, ktoré využíva nielen robotika a automatizované systémy, ale napr. i niektoré neurónové, osobitne konvolučné siete používané na rozpoznávanie obrazu v moderných systémoch umelej inteligencie.

inteligenciu. Avšak v pracovných debatách ju bolo treba nejako nazvať a tak ju nazvali umelá inteligencia (artificial intelligence).[71] I táto drobnosť vyjadruje počiatočné snahy uchopiť novú oblasť, ktorá sa s rozvojom elektronických počítačov, matematických metód a poznania v oblasti biológie javila ako síce neznáma, no veľmi lákavá.

V podkladoch, ktoré boli súčasťou žiadosti o finančnú podporu na realizáciu seminára adresovanej the Rockefeller Foundation, organizátori predstavili návrh, že „**všetky aspekty učenia, resp. akékoľvek iné črty inteligencie môžu byť v princípe tak detailne popísané, že je možné vytvoriť stroje, ktoré by inteligenciu mohli simulovať**“.[72]

Detailné predstavenie tohoto návrhu v sebe zahŕňalo celú množinu tém, ktoré mali byť na seminári diskutované: **strojové spracovanie ľudskej reči, neurónové siete, strojové učenie, abstraktné koncepty a myslenie, kreativita**... Treba povedať, že uvedené témy definujú oblasti skúmania a rozvoja umelej inteligencie dodnes.

Z dnešného uhľa pohľadu – vnímajúc problémy, s ktorými sa pri tvorbe systémov umelej inteligencie potýkajú už celé generácie odborníkov z rozličných oblastí a uvedomujúc si, že najvýkonnejšie počítače v roku 1956 boli milión krát pomalšie, než dnešné bežné chytré mobilné telefóny – je zarážajúci optimizmus, ktorý účastníci seminára ohľadom dosiahnutia umelej inteligencie mali: „myslíme si, že významný pokrok v riešení jedného alebo viacerých z týchto problémov sa dá dosiahnuť, ak starostlivo vybraná skupina vedcov bude na tom spoločne pracovať počas jedného leta“.[73]

Z dnešného uhľa pohľadu vieme, aký nereálny a idealistický pohľad účastníci seminára mali.[74] Nech sa však v nasledovných rokoch a desaťročiach dialo čokoľvek, v roku 1956

---

71 NILSSON, N. J., MCCARTHY, J. *A Biographical Memoir*. Washington D.C.: National Academy of Sciences, 2012.

72 I keď sa účastníci Dartmouthského seminára snažili diskutovať vytvorenie skutočnej inteligencie, vyjadrenie „stroje, ktoré by inteligenciu mohli simulovať“ vyjadruje jadro problému.
MCCARTHY, J. et al.: *Proposal for the Dartmouth Summer Research Project in Artificial Intelligence*. In: *AI Magazine.* [on-line]. 1955, 27(4). [cit. 3. februára 2021].
Dostupné na internete: <https://doi.org/10.1609/aimag.v27i4.1904>

73 MCCARTHY, et al.: *Proposal for the Dartmouth Summer Research Project in Artificial Intelligence*. [on-line]. [cit. 3. februára 2021].
Dostupné na internete: <https://doi.org/10.1609/aimag.v27i4.1904>

74 McCarthy na začiatku šesťdesiatych rokov zakladá Standfordský projekt umelej inteligencie „s cieľom

prakticky započal rozvoj nového vedného odboru, ktorý na Dartmouthskom seminári dostal nielen meno, ale i solídny náčrt základných cieľov.[75]

V kontexte zamerania tohoto textu je zaujímavé primárne inšpirovanie sa realizácie AI biologickými systémami a porovnávanie AI s ľudskou inteligenciou. I keď prvotný pohľad bol čiste biologicko-matematický, moderný rozvoj systémov AI a dosah ich použitia v rôznych oblastiach života priniesol veľké množstvo otázok, ktoré sa dotýkajú *koexistencie* AI a človeka, resp. spoločnosti. Ide teda o otázky psychologické, sociologické a etické. Inak povedané, **ak chceme riešiť pokročilé systémy umelej inteligencie a ich použitie, nie je možné stavať len na matematických a biologických základoch, ale musíme si uvedomiť, že akákoľvek inteligencia – ak sa má takto nazývať – tieto základy bytostne presahuje**.

1.2. Definícia pojmov

Už debata o termíne umelá inteligencia naznačuje otázky, ktoré si od čias Dartmouthského seminára bádatelia v oblasti AI znovu a znovu kladú: Koľko nám toho ešte zostáva zdolať do vytvorenia skutočne inteligentného stroja? Bude vytvorenie plnej umelej inteligencie vyžadovať reverzné inžinierstvo ľudského mozgu v celej jeho komplexite, alebo nájdeme skratku, nejakú super šikovnú množinu dnes ešte neznámych algoritmov, ktoré sa budú podieľať na vytvorení plnej umelej inteligencie? A čo je podstatné – **vieme vôbec definovať inteligenciu a čo to vlastne znamená „plná inteligencia“?**

---

vytvorenia plne inteligentného stroja v rámci dekády“.
MORAVEC, H. *Mind Children: The Future of Robot and Human Intelligence*. Cambridge, Mass.: Hardvard University Press, 1988, s. 20.
Budúci laureát Nobelovej ceny v tom istom čase predikuje: „Do dvadsať rokov budú stroje schopné vykonávať akúkoľvek prácu, ktorú môže vykonávať človek“.
SIMON, H. A. *The Shape of Automation for Men and Management*. New York: Harper&Row, 1965, s. 90.
Minského predikcia hovorila podobne nádejne: „v rámci jednej generácie budú problémy spojené s vytvorením umelej inteligencie v podstate vyriešené“.
MINSKY, M. L. *Computation: Finite and Infinite Machines*. Upper Saddle River, N.J.: Prentice Hall, 1967, s. 2.
I napriek úžasnému pokroku – osobitne v poslednej dekáde – žiadna z týchto predpovedí sa doteraz nenaplnila...

75 Štyria z účastníkov, John McCarthy, Marvin Minsky, Allen Newell a Herbert Simon sa nazývajú „veľká štvorka“ zakladateľov.

Stará latinská zásada *Qui bene distinguit, bene docet*[76] je výzvou pre kohokoľvek, kto v oblasti umelej inteligencie tvorí, alebo o nej hovorí.[77] Pretože už základné pojmy ako inteligencia, myslenie, poznanie, vedomie a city sú pre potreby bádania v oblasti umelej inteligencie nedostatočne definované. Marvin Minsky preto pre tieto pojmy razí termín „suitcase word".[78] Každý z týchto pojmov je ako veľký kufor, obsahujúci spleť rôznych významov a možností. Preto i umelá inteligencia participuje na tomto probléme a nadobúda tak rôzne významy podľa konkrétnych kontextov.

Na inteligenciu ľudského bytia nazeráme ako na schopnosť vnímať, chápať a spracovávať informácie, učiť sa, odôvodňovať, plánovať, riešiť komplexné problémy a rozhodovať. Novozélandský morálny filozof a psychológ James Flynn definuje inteligenciu ako **súbor zručností, medzi ktoré patrí abstraktné a logické myslenie, ďalej predstavivosť, a teda aj schopnosť uvažovať o hypotetických možnostiach, a tiež aj jazykový cit**.[79]

Ak hovoríme o ľudskej inteligencii, okrem kognitívnych schopností naviac **uvažujeme aj o metakognitívnych aspektoch inteligencie v rámci ktorých ľudská bytosť dokáže monitorovať, kalibrovať a prispôsobovať svoje vlastné procesy uvažovania**. Ide v zásade o uvedomovanie si svojich vlastných procesov a aktívnu schopnosť ich regulovať.[80]

Ako ľudia tak nejako podvedome vieme rozlíšiť, čo je a čo nie je inteligentné (napr. človek vs. kameň na ceste,...) a vieme takto aj inteligenciu porovnávať (napr. človek vs. šimpanz,...). Dokonca máme vytvorené aj merateľné škály pre ľudskú inteligenciu (IQ), a naviac vieme rozlišovať jej rôzne dimenzie: emocionálnu, verbálnu, logickú, sociálnu,

---

76 Kto dobre rozlišuje, dobre učí.
Podobne i Alan Turing, stojaci pri zrode AI, hovorí: „Na terminológii záleží!"

77 Súčasnosť dáva tomu za pravdu, keď sa spojenie umelá inteligencia používa ako buzzword – módny termín – zahŕňajúci všetko možné od inteligentných asistentov a expertných systémov, cez jednoduché algoritmy a aplikácie umelej inteligencie, autonómne a robotické systémy, až po sofistikované systémy typu AlphaGO, IBM Watson a ChatGPT.

78 MINSKY, M. L. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. New York: Simon & Schuster, 2006, s. 95.

79 Por. FLYNN, J. *What Is Intelligence?: Beyond the Flynn Effect*. Cambridge University Press, 2009.

80 METCALFE, J., SHIMMURA, A.P. *Metacognition: Knowing about Knowing*. MIT Press, 1994. ISBN 978-0-262-13298-5.

atď.[81]

Takže inteligenciu kategorizujeme binárne (niečo je alebo nie je inteligentné), kontinuálne (jeden objekt je inteligentnejší ako druhý) a multidimenzionálne (objekt má určitú úroveň inteligencie vo viacerých dimenziách). Pojem inteligencia teda môžeme v Minského poňatí vnímať ako *na prasknutie naplnený kufor*.

Doterajší vývoj systémov AI však ukazuje, že uvedené rozlišovanie v oblasti AI je vo veľkej miere ignorované a dôraz je skôr kladený na dva rôzne smery vývoja – vedecký a praktický.

Na strane vedeckého vývoja sa výskumníci snažia pochopiť prirodzený, t.j. biologický mechanizmus inteligencie a tento následne implementovať v počítačových systémoch.

Praktický smer vývoja sa snaží skôr jednoduchou a pragmatickou cestou vytvoriť také počítačové systémy, ktoré dokážu vykonávať úlohy rovnako alebo lepšie ako ľudia, pričom pramálo záleží na tom, či tieto programy myslia spôsobom analogickým ľudskému mysleniu.

Aktuálna neexistencia presnej definície môže byť aj výhodou pre akcelerovanie celého odvetvia umelej inteligencie v rôznych smeroch vývoja a aplikovania systémov AI. Neexistencia presnej definície AI a viac menej ignorancia základných kategorizácií sú živnou pôdou pre veľmi rôznorodý a kreatívny rozvoj tejto oblasti, keďže bádateľov vedie len hrubý zmysel pre toto smerovanie a snaha etablovať sa, či dostať sa do popredia v tejto emergentnej oblasti.[82] Súbežne sa vyvíja právne prostredie i systém štandardov a rovnako búrlivý je i rozvoj technológií a príbuzných vedných odborov, čo sa v nemalej miere podieľa na tvorbe nových riešení umelej inteligencie.

Z uvedeného vyplýva, že môže, a skutočne i existuje viacero definícií fenoménu umelej inteligencie a ich počet sa vzhľadom na neustály rozvoj týchto technológií postupne zväčšuje, resp. ich obsah sa spresňuje, alebo sa v závislosti od účelu použitia upravuje (napr. návrh definície AI v kontexte Nariadenia o umelej inteligencii z dielne EÚ).

---

81 V dobe reálneho zavádzania nástrojov AI do praxe sa začína hovoriť aj o tzv. transformačnom kvociente (TQ), t.j. o adaptabilnej schopnosti zvládať zmeny, dokonca vedieť absorbovať aj zmeny disruptívne s nedoziernymi následkami pre život, resp. pracovnú činnosť človeka.

82 MITCHELL, *Artificial Intelligence*, s. 20.
One Hundred Year Study on Artificial Intelligence. [on-line]. AI100, 2016, s. 12. [cit. 14. novembra 2021]. Dostupné na internete: <https://ai100.stanford.edu/2016-report>

Správa z roku 2016 stanfordskej *100-ročnej štúdie o umelej inteligencii*[83] definuje umelú inteligenciu ako oblasť počítačových vied, ktorá študuje vlastnosti inteligencie syntetizovaním inteligencie.[84]

100-ročná štúdia o umelej inteligencii (AI100) v správe z roku 2021 dopĺňa alternatívnu definíciu AI: **umelá inteligencia je o vytváraní systému, ktorý vykazuje také správanie, o ktorom si myslíme, že vyžaduje inteligenciu**.[85] Myslíme si, že s uvedenou definíciou môžeme pokračovať v našej ďalšej rozprave.

Stále však nesmieme zabúdať na základný rozdiel medzi súčasnou umelou a ľudskou inteligenciou, t.j. **rozdiel medzi funkcionalistickou umelou inteligenciou a komplexnou inteligenciou ľudskej bytosti**!

1.3. Základné vlastnosti a delenie systémov umelej inteligencie

Na základe doterajšej debaty ohľadom definície a kategorizácie inteligencie – systémy, ktoré by boli schopné vykazovať inteligentné správanie, musia spĺňať dve základné kritériá:

- **autonómnosť** – schopnosť samostatne konať, t.j. schopnosť systému vykonávať úlohy v komplexnom prostredí bez neustáleho vedenia používateľom.[86]
- **adaptabilnosť** – schopnosť sa prispôsobovať, t.j. vlastnosť systému zlepšovať svoj výkon (a schopnosti) učením sa zo skúseností.

Z predchádzajúcej kapitoly si pripomeňme, že o inteligencii môžeme hovoriť v kategóriách kontinuity (jeden objekt je inteligentnejší ako druhý) a viacerých dimenzií (inteligencia v rôznych oblastiach).

83 Stanfordská 100-ročná štúdia o umelej inteligencii (AI100) je projekt, ktorý sa začal v decembri 2014 s ambíciou pravidelne vyhodnocovať AI panelom odborníkov raz za 5 rokov. Ide o aktivitu zameranú na pravidelnú analýzu súčasného stavu rozvoja umelej inteligencie a jej vplyvu na ľudí.

84 SIMON, H. A. *Artificial Intelligence: An Empirical Science*. In *Artificial Intelligence* 77. 1995, č. 2, s. 95-127.
One Hundred Year Study on Artificial Intelligence. [on-line]. AI100, 2016, s. 13. [cit. 14. novembra 2021]. Dostupné na internete: <https://ai100.stanford.edu/2016-report>

85 *One Hundred Year Study on Artificial Intelligence*. [on-line]. AI100, 2021, s. 78. [cit. 15. novembra 2021]. Dostupné na internete: <https://ai100.stanford.edu/2021-report>

86 Autonómnosť nemôžeme zamieňať s automatizáciou, pri ktorej ide len o vykonávanie činností bez, resp. s minimálnym zásahom človeka. Automatizované systémy schopnosť samostatne konať nemajú.

Preto i systémy môžu byť adaptívne a autonómne len v určitej oblasti, t.j. sú schopné riešiť určité úlohy „inteligentným spôsobom“, pričom v ostatných oblastiach zlyhávajú. Takéto systémy nazývame **úzko špecializované umelé inteligencie (narrow AI)**[87]. Ide teda o vysoko špecializované systémy, ktoré sú optimalizované na zvládnutie konkrétnej úlohy, resp. množiny úloh.[88]

Všeobecná umelá inteligencia (general AI)[89] je systém, ktorý dokáže zvládnuť akúkoľvek intelektuálnu úlohu. V podstate ide o umelú inteligenciu, ktorá je na úrovni človeka.[90]

Všeobecná umelá inteligencia – i napriek bombastickým titulkom v novinách a niektorým futurologickým predpovediam – patrí v súčasnosti do oblasti sci-fi.[91]

Analogicky rozlišujeme umelú inteligenciu ako *silnú* a *slabú* na základe rozlišovania medzi inteligentnými systémami a systémami, ktoré inteligentne konajú, pričom inteligenciu len simulujú.[92]

Silná umelá inteligencia (strong AI) je skutočne inteligentná (a potencionálne i „uvedomelá“), t.j. skutočne aj rozumie tomu, či rieši a vykonáva. Skutočne inteligentná AI je súčasne aj všeobecnou, pretože má schopnosť generalizovať, t.j. zovšeobecňovať a prenášať, či adaptovať naučené schopnosti na iné úlohy (čo mimochodom patrí k základom ľudského myslenia).

87 Používa sa skratka **ANI – Artificial Narrow Intelligence.**

88 Napríklad IBM Deep Blue dokáže poraziť najlepších šachistov sveta a AlphaGo Lee, resp. Zero už dlhodobo deklasuje najlepších hráčov hry Go, no ak by tieto systémy mali byť použité v autonómnych vozidlách, nedokázali by sa ani rozbehnúť na rovnej ceste.

89 Používa sa skratka **AGI – Artificial General Intelligence**.

90 Niekedy sa z okruhu AGI samostatne vyčleňuje ešte **ASI – Artificial Super Intelligence**, t.j. umelá inteligencia, ktorá by bola naprieč všetkými oblasťami oveľa inteligentnejšia ako človek. Úvahy o ASI sa mnohokrát spájajú s konceptom *singularity v oblasti AI*, v ktorej umelá inteligencia so schopnosťou učiť sa, zlepšovať sa a samostatne sa vyvíjať rýchlo dosiahne a následne extrémne prevýši inteligenciu človeka. Viac o vízii Raya Kurzweila a o singularite je uvedené v poznámke č. 16.

91 Vieme vytvoriť inteligentné systémy, ktoré v konkrétnych oblastiach sú lepšie než ľudia, no nedokážeme realizovať systémy, ktoré to dokážu všeobecne.

92 Tento rozdiel akcentuje zvýraznený odstavec a poznámky o základnom probléme v kapitole 1.1.: „Základným problémom – a jeho podstata sa týka prakticky takmer všetkých systémov umelej inteligencie dneška – je skutočnosť, že i keď nejaký systém dokáže odpovedať tak, akoby konverzácii rozumel, neznamená to, že je inteligentný a že konverzácii aj skutočne rozumie.“

Slabá umelá inteligencia (weak AI) vykazuje inteligentné správanie na základe modelov a aplikovaných metód i dát, na ktorých sa učí (je natrénovaná). Ide teda o systémy, ktoré sú zamerané na riešenie konkrétnych úloh a sú závislé na ľudskom vstupe a konfigurácii.[93] Slabá AI reprezentuje systémy, ktoré aktuálne máme, t.j. počítačové systémy, ktoré vykazujú inteligentné správanie.

V súčasnosti skratkou **ANI** vyjadrujeme úzku a slabú umelú inteligenciu (narrow and weak AI) a pod skratkou **AGI** zvykneme rozumieť všeobecnú a silnú umelú inteligenciu (general and strong AI).

Naviac – v kontexte aktuálnych pokrokov v oblasti generatívnych systémov AI – sa stále viac hovorí o rôznych stupňoch AGI s viac menej otvorenou snahou radiť tie najlepšie jazykové, resp. multimodálne modely na pomedzie medzi ANI a AGI. Delenie na ANI a AGI nestačí, a preto sa zvykne AGI deliť do štyroch fáz (stupňov):[94]

- BAGI (below-human) – systémy približujúce sa schopnostiam ľudskej inteligencie
- HAGI/HLAGI (human-level AGI) – systémy na úrovni človeka
- MAGI (moderately-superhuman AGI) – systémy presahujúce schopnosti človeka
- SAGI/ASI (superintelligent AGI) – super inteligencia nezávislá na ľudskej infraštruktúre a smerujúca, resp. realizujúca singularitu v oblasti AI.

Všetky metódy a systémy umelej inteligencie, ktoré dnes používame, spadajú do kategórie špecializovaných systémov AI (narrow AI), pričom súčasný rozvoj v tejto oblasti napreduje míľovými krokmi, takže viacerí odborníci hovoria o najlepších súčasných systémoch ako o BAGI (treba dodať, že približne rovnako veľká skupina odborníkov s týmto názorom nesúhlasí:-).[95]

93 Vytvorenie skutočne dobre fungujúceho systému AI vyžaduje skutočných odborníkov, schopných aplikovať správne metódy i správne nakonfigurovať a parametrizovať daný systém. Pre neznalých to môže vyzerať ako mágia, či voodoo:-)
Por. MITCHELL, *Artificial Intelligence*, s. 98.

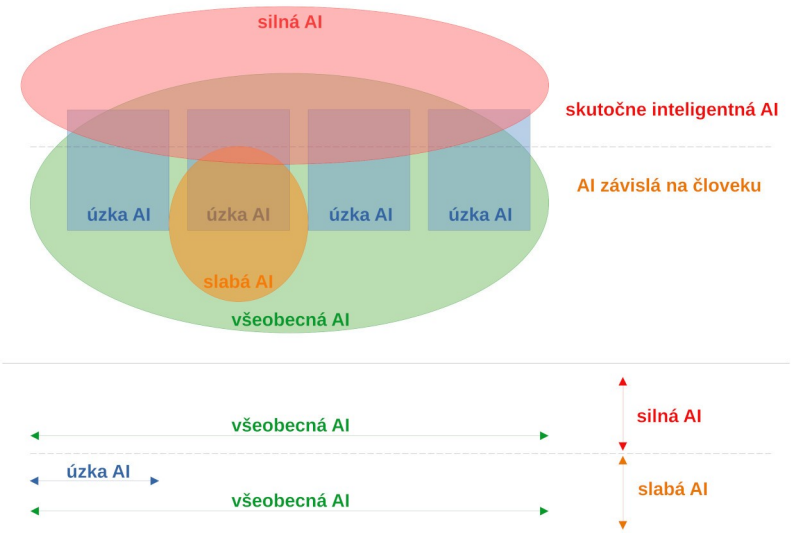
94 GABE, M. *Four Phases of AGI*. [on-line]. [cit. 22. december 2024].
Dostupné na internete: <https://www.lesswrong.com/posts/qeJomTN2yp5tQG4rL/four-phases-of-agi>

95 Na platforme X je veľmi známe štipľavé, no zároveň odborné odmietanie radenia súčasných generatívnych systémov do kategórie BAGI od známeho profesora Gary Marcusa, kognitívneho vedca a jedného z autorov pokročilých testov inteligenčných schopností systémov AI.
Gary Marcus. [on-line]. [cit. 28. december 2024].
Dostupné na internete: <https://x.com/garymarcus>

Príkladom úzkej a slabej umelej inteligencie môže byť algoritmus pre hranie šachu alebo hry Go, autonómne riadenie vozidla, systém na automatické odporúčanie obsahu v rámci služieb TikTok alebo Youtube a pod. Vždy ide o systémy optimalizované na konkrétnu úlohu, pričom pracujú na základe modelov a natrénovania, ktoré realizovali ľudské tímy.

Keďže všetky súčasné systémy umelej inteligencie sú tzv. ANI, t.j. patria do kategórie úzkej a slabej AI (prípadne BAGI), príklady všeobecnej a silnej umelej inteligencie poznáme len z teoretických návrhov alebo zo science-fiction. Medzi Sci-Fi predstavy patrí napríklad postava Terminátora, Holly zo seriálu Červený trpaslík, počítač HAL 9000 z románu Arthura C. Clarka 2001: Vesmírna odysea alebo viacero AI protagonistov z diel Isaaca Asimova.

Nasledujúci obrázok vyjadruje dve znázornenia vzájomného vzťahu a pomeru medzi silnou a slabou, všeobecnou a úzko špecializovanou umelou inteligenciou.



Obr. č. 1. Vzťah medzi silnou a slabou, všeobecnou a úzko špecializovanou AI.

1.4. Symbolické a subsymbolické systémy ako súčasť anarchie metód

Ak sme v kapitole 1.1. o úsvite umelej inteligencie a osobitne o Dartmouthskom seminári trošku medzi riadkami čítali, môžeme tušiť, že už od počiatkov výskumu existoval nespočet pohľadov na to, akým spôsobom treba uchopiť problematiku umelej inteligencie a akými cestami by sa jej výskum a vývoj mal uberať. Ani nasledovné desaťročia na rozsahu tejto spletitosti neubrali. A preto, vzhľadom na našu nedostatočnosť chápania

inteligencie ako takej[96] a z nej prameniacej neschopnosti uchopiť vytvorenie všeobecnej umelej inteligencie, miesto jasnej cesty vývoja AI sa vývoj a následný pokrok dosahuje v takej rozmanitosti a spletitosti metód AI, že môžeme hovoriť o anarchii metód.[97]

Výskum umelej inteligencie tak obsahuje širokú množinu prístupov, ktorá – na prvý pohľad – môže byť pre vývoj systémov AI brzdou. Máme však za to, že v kontexte našej nedostatočnosti chápania inteligencie ako takej, je tento prístup asi jediný možný. Na jednej strane sú tak preskúmané smery a metódy, ktoré k cieľu nevedú, na strane druhej súčasné úspechy v oblasti rozvoja AI stavajú na schopnosti nielen nachádzať postupy a algoritmy nové, ale i kombinovať rôzne metódy a prístupy,[98] prípadne využiť staršie, javiace sa ako neúspešné, avšak zdokonalené natoľko, že prinášajú kýžené ovocie.[99]

I napriek uvedenej anarchii však môžeme aspoň principiálne – filozoficky rozdeliť túto širokú množinu metód do dvoch za nimi stojacich prístupov:

- symbolická AI
- subsymbolická AI

Symbolická umelá inteligencia ide cestou vytvárania umelej inteligencie na báze ľudského myslenia, t.j. pojmov, slov, fráz (= symboly) a vzťahov medzi nimi. Symbolické systémy na základe definovaných pravidiel a postupov („ak niečo, tak potom toto“) môžu jednotlivé symboly spracovávať a vykonávať priradené úlohy.[100] Pre symbolickú AI sa význam jednotlivých symbolov odvíja od spôsobu ich kombinácie, vzájomných vzťahov

96 Pojednávali sme o tom v kapitole 1.2.

97 LEHMAN, J., CLUNE, J., RISI, S.: *An Anarchy of Methods: Current Trends.* In: *How Intelligence Is Abstracted in AI*. IEEE Intelligent Systems. 2014, 29, č. 6, s. 56-62.

98 Napr. realizácia konvolučných sietí.

99 Jeden konkrétny príklad z praxe: „The return of patch-based self-supervision! It never worked well and you had to bend over backwards with ResNets (I tried). Now with ViT, very simple patch-based self-supervised pre-training rocks! First BeIT, now Masked AutoEncoders i1k=87.8%“ [on-line]. Twitter. [cit. 18. novembra 2021].
Dostupné na internete: <https://twitter.com/giffmana/status/1459092079020285976>
HE, K., CHEN, X., XIE, S., LI, Y., DOLLÁR, P., GIRSHICK, R.: *Masked Autoencoders Are Scalable Vision Learners* [on-line]. Facebook AI Research (FAIR). [cit. 18. novembra 2021].
Dostupné na internete: <https://arxiv.org/abs/2111.06377>

100MITCHELL, *Artificial Intelligence*, s. 21.

a operácií, ktoré môžu byť nad nimi vykonávané.[101]

Keďže **logika je nutnou podmienkou nášho chápania všeobecnej inteligencie**, symbolické systémy sa snažia logickými operáciami riešiť rôznorodé úlohy vyžadujúce nejaký stupeň inteligencie. Striktná formálnosť logického uvažovania umožňuje algoritmizovať ho a previesť do strojovej formy.

Počas prvých dvoch či troch dekád výskumu umelej inteligencie symbolické systémy prevládali. Jednou z prvých implementácií symbolickej AI bol General Problem Solver (GPS)[102], ktorý vedel riešiť také problémy, ako napr. hlavolam Misionári a kanibali[103]. Jedným z vrcholov symbolických systémov sú expertné systémy,[104] ktoré využívali (a dodnes využívajú) človekom vytvorené pravidlá pre riešenie úloh spojených s lekárskou diagnostikou či právnymi rozhodnutiami, alebo zadania v spravodajských službách[105] atď.

Prvé symbolické systémy využívali tzv. výrokovú logiku,[106] no jej aplikácia v systémoch AI rýchlo narazila na limity a obmedzenia pri snahe logicky popísať všetky možnosti, ktoré pri vykonávaní logických operácií nad symbolmi reálneho sveta môžu nastať. Jednoducho povedané, výrokovou logikou nedokážeme popísať tieto úlohy v dostatočne všeobecnej rovine a algoritmy umelej inteligencie s výrokovou logikou nedokážu zovšeobecňovať bez predchádzajúceho masívneho natrénovania, resp. popisu všetkých možných situácií.[107]

---

101 MITCHELL, *Artificial Intelligence*, s. 23.

102 NEWELL, A., SIMON, H. A. *GPS: A Program That Simulates Human Thought*. Santa Monica, California: P-2257, Rand Corporation, 1961.

103 *Missionaries and cannibals problem* [on-line]. [cit. 18. novembra 2021]. Dostupné na internete: <https://en.wikipedia.org/wiki/Missionaries_and_cannibals_problem>

104 I keď pre autora už v tých časoch boli neurónové siete a subsymbolická AI oveľa príťažlivejšia, na konci deväťdesiatych rokov bol súčasťou tímov, ktoré pre reálne použitie zvažovali nasadenie expertných systémov ako v tom čase jedinej reálnej voľby.

105 Využívajú sa napr. systémy na inteligentné vyhľadávanie a hlbokú analýzu dát (data mining) alebo na rozpoznávanie objektov (supervector machine).

106 Výroková logika sa zaoberá výrokmi, ich pravdivosťou a odvodzovaním. Pojem výrok je možné chápať ako tvrdenie v tvare oznamovacej vety, ktoré môže byť pravdivé alebo nie. *Výroková logika - základne pojmy* [on-line]. [cit. 15. januára 2024]. Dostupné na internete: <https://kurzy.kpi.fei.tuke.sk/lpi/labs/01.html>

107 Analogický problém majú aj na štatistike a pravdepodobnosti postavené metódy s porovnateľnými vyjadrovacími schopnosťami.

Oveľa sľubnejšie sa javia systémy využívajúce logiku prvého rádu, ktoré ponúkajú formálny jazyk schopný širších a všeobecnejších formulácií pre formulovanie logických operácií a tým aj rôznych aspektov inteligencie. **Kým výroková logika stavia na výrokoch, ktoré môžu byť pravdivé alebo nepravdivé, logika prvého rádu stavia na rôznorodých vzťahoch medzi objektami.**[108] **V rámci riešenia úloh, ktoré boli založené na presných a istých informáciách, systémy využívajúce logiku prvého rádu dosiahli veľkú dokonalosť**.

Pri snahe riešiť zložitejšie úlohy, ktoré mali napr. nepresné vstupy, či vyžadovali rozpoznávanie objektov v reálnych obrazových scénach s vizuálnym šumom a pod., však symbolické systémy zlyhávali. **Nádejné AI systémy postavené na logike prvého rádu tak pohoreli na neschopnosti zvládnuť neisté a nejednoznačné informácie**.[109]

Popritom treba uviesť, že symbolické systémy umelej inteligencie stále existujú, v niektorých oblastiach AI sa stále využívajú a ovocie ich doterajšieho vývoja, resp. kombinácia s rôznymi algoritmami subsymbolických metód je zvažovaná ako reálny vklad do diskusie o vývoji nových systémov ANI a AGI.[110]

Subsymbolická umelá inteligencia a jej rozvoj boli inšpirované pokrokom v neurovede. Subsymbolický prístup k AI sa snaží uchopiť naše myšlienkové procesy, ktoré by sme mohli nazvať niekedy nevedomými, či automatickými, a ktoré sú základom tzv. rýchleho vnímania (fast perception), čo využívame napr. pri rozpoznávaní tvárí alebo identifikácii hovorených slov.

Subsymbolické programy umelej inteligencie tak neobsahujú súbor presných softvérových postupov na úrovni logického myslenia, ale sú tvorené len stohom rovníc, pre nezainteresovaného len húštinou ťažko interpretovateľných operácií s číslami.

Por. RUSSELL, *Human Compatible*, s. 270.

108 Napríklad v hre Go by sme v rámci výrokovej logiky museli vytvoriť neskutočne veľa pravidiel prakticky pre všetky možné ťahy a polohy jednotlivých kameňov. Pomocou logiky prvého rádu vieme pravidlá definovať všeobecne a veľmi jednoducho: pre všetky časové kroky T a pre všetky polohy P a pre všetky farby C, ak je C na ťahu v čase T a P je v čase T neobsadené, potom je pre C legálne posunúť kameň na pozíciu P v čase T. Ak k tomu pridáme niekoľko ďalších definícií, napr. definovanie miest a dvoch farieb na šachovnici, alebo čo to znamená neobsadená pozícia, máme základ úplných pravidiel hry Go. Pravidlá zaberajú v logike prvého rádu približne toľko miesta ako v angličtine alebo v slovenčine.

109 Por. RUSSELL, *Human Compatible*, s. 271.

110 Por. MITCHELL, *Artificial Intelligence*, s. 24.

Subsymbolické systému sú navrhnuté tak, aby sa učili vykonávať úlohy na základe dát.[111]

Symbolická AI – veľmi zjednodušene povedané – sa pomocou matematickej logiky snaží emulovať procesy myslenia[112]**, subsymbolická AI – znovu zjednodušujúc – sa v mnohých prípadoch usiluje o emuláciu činnosti mozgu na úrovni neurónov.**

Symbolický prístup – využívajúci prísnu logiku a stavajúci na jasne definovaných charakteristikách prostredia a vzťahov medzi objektami činnosti systémov AI – dosahoval svoj vrchol v osemdesiatych rokoch minulého storočia.[113] Tento prístup však dokázal byť úspešný len pri riešení špecifických typov problémov založených na konkrétnych charakteristikách prostredia, vybraných interakciách systémov AI s týmto prostredím a zadanej konkrétnej množine cieľov. Všetko ostatné, čo by mohlo a malo byť predmetom činnosti akejkoľvek inteligencie, však bolo pre tieto systémy tabu – symbolické systémy zlyhávali a zlyhávajú pri riešení problémov, ktoré sa nedajú exaktne opísať, a v reálnych prostrediach, ktoré nie je možné deterministicky uchopiť a sú plné *nejasných* informácií.[114]

Preto **väčšina moderných implementácií systémov umelej inteligencie (medzi ktoré patria algoritmy strojového učenia vo všeobecnosti a osobitne neurónové siete) vychádza zo subsymbolického prístupu, ktorý sa snaží tieto problémy uchopiť a v určitej miere aj úspešne riešiť** – či už klasickými metódami, ako sú napríklad regresné a štatistické metódy, alebo emuláciou činnosti mozgu na úrovni neurónov prostredníctvom tzv. neurónových sietí.

Princíp neurónových sietí síce poznáme už osemdesiat rokov, keďže prvý matematický model týchto sietí navrhli v roku 1943 Warren McCulloch a Walter Pitts ako ovocie

111 Por. MITCHELL, *Artificial Intelligence*, s. 24.

112 Procesy, ktoré u ľudí zahŕňajú zmyslové vnímanie, abstrahovanie pojmov, vytváranie súdov a úsudkov.

113 Išlo o také systémy ako Fifth Generation project japonskej vlády, projekty vládnej agentúry DARPA v USA, prípadne britská Strategic Computing Initiative, ktoré využívali tzv. first-order logic pretavenú do symbolického programovania v programovacom jazyku Prolog.
Por. RUSSELL, *Human Compatible*, s. 271.

114 Hovoríme o riešení stochastických procesov a pre ne vytvorených modelov, keďže obsahujú prvok náhodnosti alebo neistoty.
RUSSELL, S., NORVIG, P. *Artificial Intelligence: A Modern Approach, 3rd edition*. Pearson, 2009, s. 177. ISBN: 978-9-332-54351-5.

skúmania fungovania biologických neurónov,[115] ale skutočný rozmach vo využívaní týchto sietí prichádza v porovnaní so symbolickými systémami AI oveľa neskôr. Dôvodov je viacero: jednoduchší návrh a nasadenie symbolických systémov (výrokovú logiku a logiku prvého rádu je možné výborne algoritmizovať a programovať), počiatočná neexistencia kľúčových prvkov realizácie neurónových sietí (napr. back-propagation algorithm – algoritmus spätného šírenia chyby pre realizovanie spätnej väzby vo viacerých vrstvách) a následne oneskorený vývoj adekvátnych modelov týchto sietí, neexistencia skutočne rozsiahlych datasetov (množín štruktúrovaných i neštruktúrovaných dát) a nedostatočná kapacita výpočtových systémov. Vyriešenie uvedených problémov znamenalo nielen rýchle rozšírenie neurónových sietí, ale aj veľký nárast v ich výkone a schopnostiach.

1.5. Systémy inšpirované činnosťou mozgu na úrovni neurónov

Ako sme v predchádzajúcej kapitole uviedli, **väčšina moderných implementácií systémov umelej inteligencie vychádza zo subsymbolického prístupu, ktorého jednou z inšpirácií je činnosť mozgu na úrovni neurónov**.

Tento prístup si môžeme objasniť na jednom z prvých príkladov subsymbolického, mozgom inšpirovaného programu AI, ktorý bol na sklonku 50-tych rokov minulého storočia pod názvom **perceptron** vyvinutý psychológom Frankom Rosenblattom.[116]

I keď by sa mohlo zdať, že v dnešnej dobe ide z pohľadu počítačovej vedy o príklad z praveku, perceptron bol nielen míľnikom v oblasti umelej inteligencie, ale predovšetkým principiálne ovplyvnil vývoj subsymbolických systémov a bol starým otcom moderných a úspešných súčasných systémov AI, ktoré poznáme ako hlboké neurónové siete.

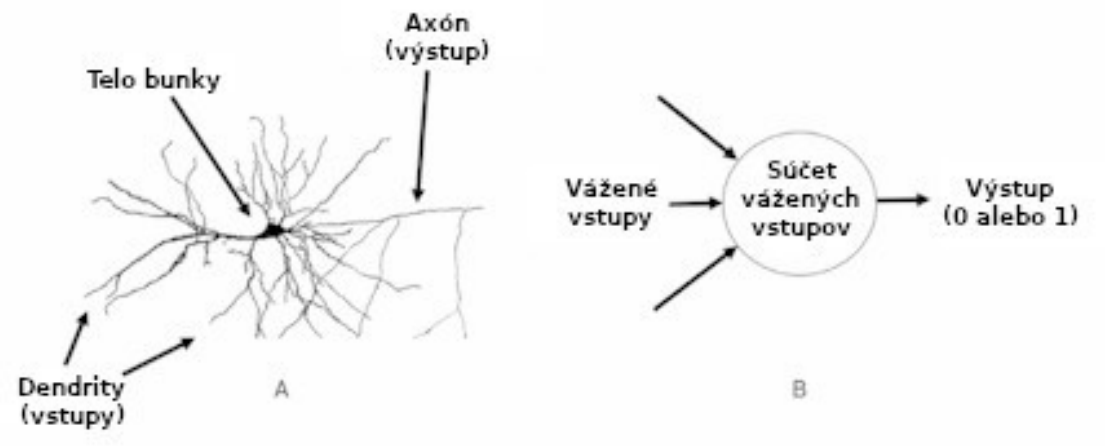
Fungovanie perceptronu bolo principiálne inšpirované spôsobom, ako neuróny spracúvajú informácie: neurón ako mozgová bunka spracúva elektrické, resp. chemické vstupy (vzruchy) z iných neurónov, s ktorými je prepojená. Veľmi zjednodušene povedané, neurón zráta všetky vstupy z ostatných neurónov a ak celková hodnota presiahne určitú hraničnú hodnotu, neurón vyšle impulz. Pritom je dôležité, že jednotlivé spojenia od ostatných neurónov (synapsy) sú rozlične silné. Keď následne neurón ráta výsledný sumár, vstupy

115 MCCULLOCH, W.S., PITTS, W. *A Logical Calculus of the Ideas Immanent in Nervous Activity*. Bulletin of Mathematical Biophysics, 1943, Vol. 5, s. 115-133.

116 ROSENBLATT, F. *The Perceptron: A Probabilistic Model of Information Storage and Organization in the Brain*. In: *Psychological Review.* 1958, 65, č. 6, s. 386-408.

zo silnejších synáps majú väčšiu váhu než vstupy zo synáps slabších. Podľa neurovedcov práve úprava sily jednotlivých spojení medzi neurónmi je jedným z kľúčových aspektov mechanizmu učenia sa mozgu.

Perceptron je v podstate počítačovou simuláciou vyššie uvedeného procesu spracovávania informácií v neurónoch. Túto simuláciu vyjadruje obrázok č. 2., pričom časť A zobrazuje neurón s označenými rozvetvenými vláknami, ktoré privádzajú vstupy do bunky (dendrity) a výstupným kanálom (axón). Časť B je schematickým náčrtom jednoduchého perceptronu, ktorý sčítava vstupy a v prípade, že výsledný súčet je rovný alebo väčší ako nastavená prahová hodnota, výstup je 1 (analogické vyslaniu impulzu u neurónu), v opačnom prípade je hodnota výstupu 0.



Obr. č. 2: A – neurón v mozgu, B – jednoduchý perceptron.[117]

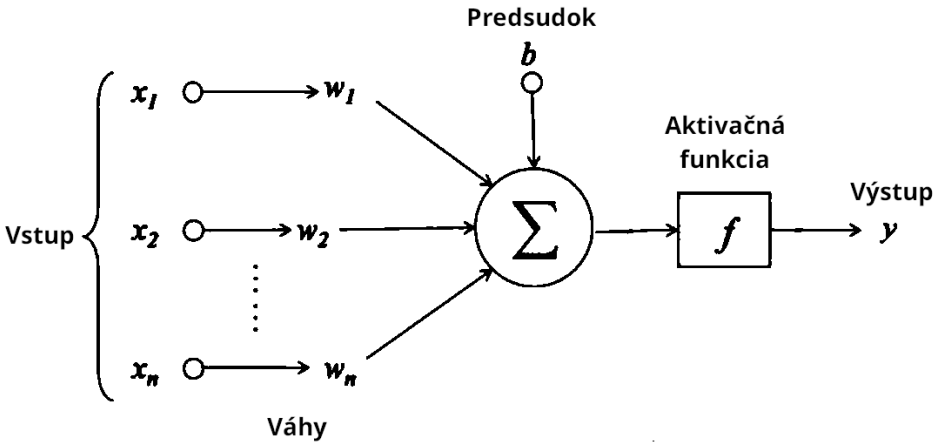
V schematickom náčrte perceptronu, ako simulácie neurónovej bunky, je slovným spojením „vážené vstupy“ vyjadrený podstatný aspekt, bez ktorého nie je možné simulovať mechanizmus učenia sa mozgu – a to vyjadrenie sily jednotlivých vstupných dendritov a ich úprava ako súčasť tohoto mechanizmu učenia sa, t.j. vyjadrenie váhy jednotlivých vstupov perceptronu a proces ich modifikácie ako súčasť učenia sa (adaptácie) systému umelej inteligencie.

Možno si ani neuvedomujeme, ako tento jednoduchý koncept využívame pri niektorých svojich rozhodnutiach. Čo nás napríklad môže viesť k zhliadnutiu konkrétneho filmu v kine? Referencie od priateľov, ktorí tento film už videli, pričom ich názor určite nedávame na rovnakú úroveň – úsudku a vkusu konkrétneho priateľa dôverujeme viac, inému menej. Vytvára sa nám tak množina vážených vstupov, pri ktorých ak pozitívne naladenie

117 MITCHELL, *Artificial Intelligence*, s. 25, upravené autorom.

pre daný film presiahne určitú úroveň, rozhodneme sa pre jeho zhliadnutie v kine, či nákup v on-line službe. Samozrejme, prichádzajú do úvahy aj ďalšie vplyvy (tzv. *hyperparametre*), ako napr. reklama na film (*prahová* konštanta, resp. *skreslenie*), konkrétne preferencie pre filmový žáner a hercov (variabilita prahovej konštanty, t.j. úrovne, ktorá ak je presiahnutá, rozhodneme sa pre zhliadnutie či nákup filmu) a pod.[118] A pri výbere ďalšieho filmu na zhliadnutie úplne podvedome prichádza k zmenám v dôvere k úsudku jednotlivých našich priateľov – na základe toho, do akej miery bola ich rada pri predchádzajúcom filme relevantná a korešpondujúca s dojmom, ktorý v nás zhliadnutý film zanechal (automatický mechanizmus na zmenu váhy jednotlivých „vstupov“).

Analogicky k uvedenému príkladu bol jednoduchý koncept perceptronu postupne rozvinutý do podoby, v ktorej sú v systémoch AI implementované samoučiace sa mechanizmy na zmenu váhy vstupov, aplikované prahové konštanty (b) k súčtu vážených vstupov, resp. postupne vyvíjané nové generácie modelov neurónov od McCullochových Pittsových neurónov s prahovými, resp. sigmoidálnymi hradlami až po neuromorfné architektúry postavené na tzv. spike neurónoch (a celých spike neurónových sieťach), prípadne na neurogenetickom modelovaní.[119] Základné prvky klasických umelých neurónov sú uvedené na obr. č. 3.[120]



Obr. č. 3: Základné prvky umelých neurónov.

Na rozdiel od systémov symbolickej AI, ktorej procesy „inteligencie“ boli výsledkom presných pravidiel a vzťahov medzi symbolmi (pojmy, frázy, slová,...),

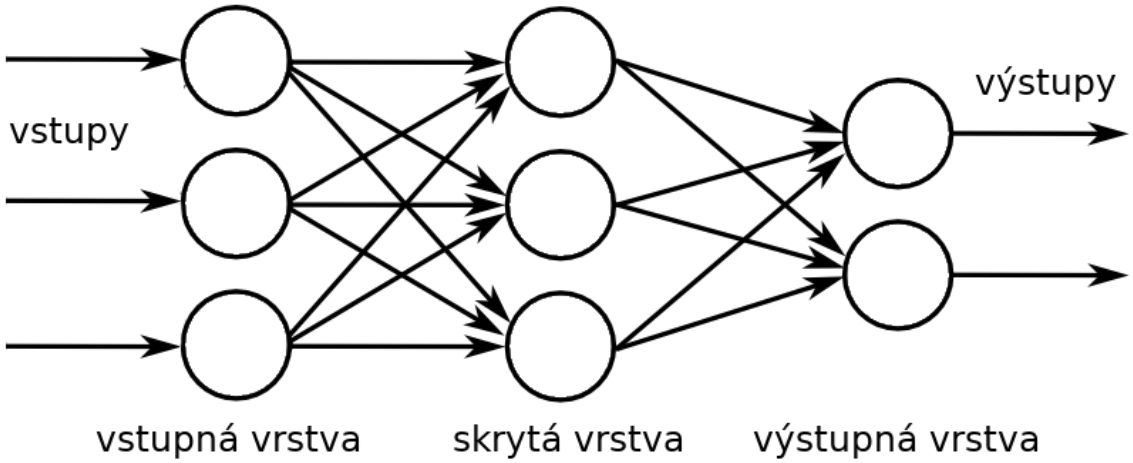
118 Inšpirovali sme sa príkladom prof. Melanie Mitchell, ktorý sme rozviedli o ďalšie aspekty systémov AI. Por. MITCHELL, *Artificial Intelligence*, s. 25.

119 Tejto problematike sa aspoň okrajovo venujeme v kapitole 1.7.

120 Medzi najpoužívanejšie aktivačné funkcie patrí Sigmoid, ReLU, Linear, Tanh a následne Softmax.

u subsymbolických systémov je všetka ich „inteligencia“ zakódovaná v číslach – parametroch, ktoré reprezentujú váhy a prahy.[121] Ide teda o stoh rovníc (viď predchádzajúca kapitola), ktoré fungujú na základe správneho nastavenia váh a prahových hodnôt. Tento proces môžeme nazvať učením, pričom systémy sa učia na vzorkách učebných (trénovacích) dát.

Realizácia systémov umelej inteligencie na základe automatického (samostatného) učenia sa z existujúcich vzoriek, dát, prípadne skúseností[122] sa nazýva strojovým učením (machine learning). Výsledkom je nachádzanie vzorov vo vstupných dátach a také nastavenie váh a prahových hodnôt, ktoré systému umožní poskytovať relevantné výsledky a rozhodnutia, to znamená adekvátne reagovať na rôzne vstupné hodnoty bez toho, aby bol na ne explicitne naprogramovaný, iba na základe informácií, ktoré sa už naučil.[123] [124]



Obr. č. 4: Viacvrstvová neurónová sieť.[125]

121 Základom pre moderné neurónové siete boli tzv. konekcionistické siete (usporiadanie a prepojenie viacerých „neurónov“), v ktorých termín *connectionist* vyjadruje ideu, že „poznanie“ v týchto sieťach sa nachádza vo vážených spojeniach medzi jednotlivými uzlami.
RUMELHART, D. E., MCCLELLAND, J. L. *Parallel Distributed Processing.* Bradford Book, 1986, zv. 1/2.

122 Pod skúsenosťou môžeme rozumieť napr. dáta zo senzorov robotického systému, ktorý práve narazil na prekážku a pod.

123 **Váhy spojení (synáps) sa v procese učenia menia, až kým sa nedosiahne spoľahlivá zhoda medzi vzorcom aktivity na vstupnej a výstupnej vrstve**. Napr. správna detekcia písaného textu, označenie prekážky na ceste, rozpoznanie hovoreného slova,...

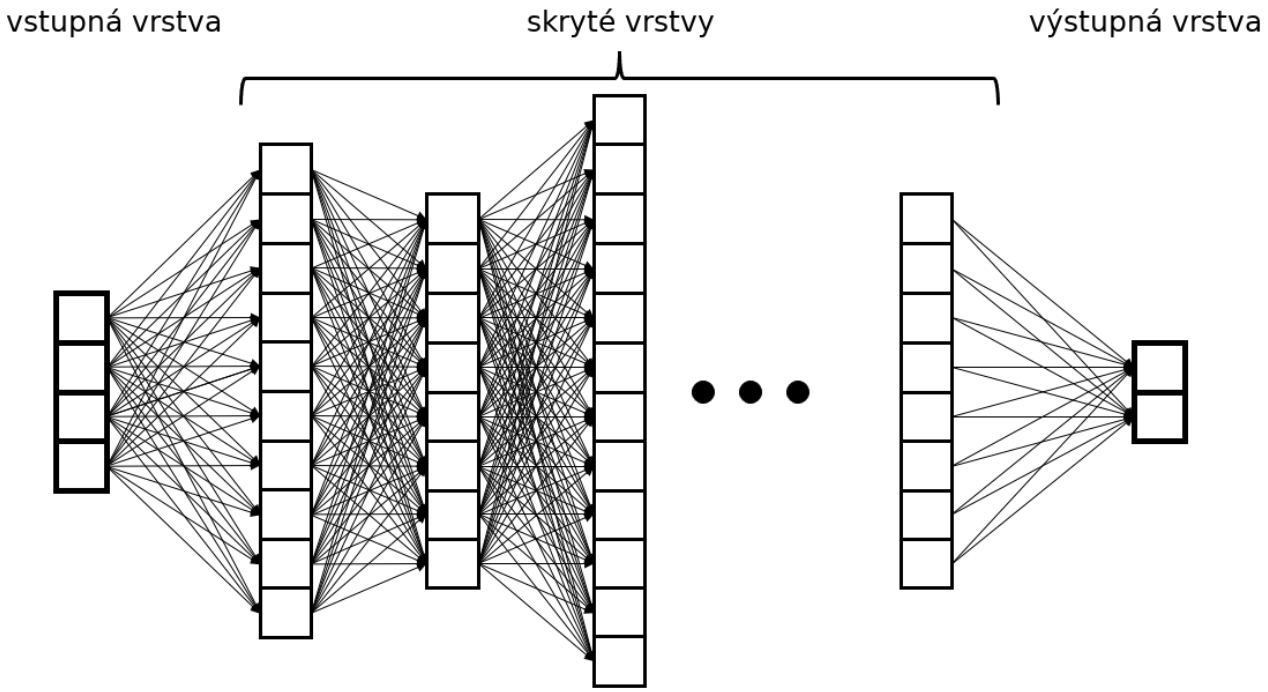
124 Por. *Algoritmy strojového učenia I.* [on-line]. [cit. 3. januára 2022].
Dostupné na internete: <https://umelainteligencia.sk/algoritmy-strojoveho-ucenia/>

125 Wikimedia.org, licencia CC, upravené autorom.

Analogicky k perceptronu, **simulácia viacerých poprepájaných neurónov tvorí neurónovú sieť**, pričom jednotlivé simulácie neurónov – uzly (node/unit) sú radené do vrstiev (layer). Vstupné dáta sú postupne spracúvané jednotlivými – skrytými vrstvami (hidden layers), ktoré obsahujú skryté uzly. Skryté vrstvy môže predchádzať ešte samostatná vstupná vrstva (input layer). Posledná vrstva, ktorej výstupom je výsledok činnosti neurónovej siete, sa nazýva výstupná (output layer). Uzly medzi jednotlivými vrstvami sú navzájom poprepájané váženými spojeniami.

Ak má neurónová sieť viac než jednu vrstvu, nazýva sa viacvrstvová (multilayered network). Jej príklad je uvedený na obr. č. 4.

Sieť, ktorá obsahuje viac ako jednu skrytú vrstvu, sa nazýva **hlbokou neurónovou sieťou** (deep neural network, prípadne len deep network), ako je uvedené na obr. č. 5.



Obr. č. 5: Hlboká neurónová sieť.[126]

Súčasné sofistikované a najvýkonnejšie systémy umelej inteligencie využívajú typ strojového učenia, ktorý sa nazýva **hlboké učenie** (deep learning). Algoritmy hlbokého učenia sa učia trénovaním na obrovskom množstve dát prostredníctvom skrytých vrstiev prepojených uzlov, ktoré – ako sme uviedli – sa označujú ako hlboké neurónové siete. **Keďže hlboké neurónové siete majú veľa vrstiev, umožňujú realizovať zložitejšie**

126 Wikimedia.org, licencia CC, upravené autorom.

transformácie a riešiť zložitejšie problémy. Každá vrstva, resp. skupina vrstiev môže mať (a obvykle aj má) svoju špecifickú úlohu. Takto je možné realizovať pomerne zložité procesy, nie nepodobné procesom prebiehajúcim v ľudskom mozgu, ktorý má tiež veľa vrstiev poprepájaných medzi sebou.[127]

1.5.1. „Ahoj, svet“ v oblasti neurónových sietí

Pri výučbe programovania a spoznávaní konkrétnych programovacích jazykov býva jedným z prvých programov, ktoré študent vytvorí, tzv. „Hello, world!“ („Ahoj, svet!“). V mnohých učebniciach programovania sa používa ako príklad jedného z prvých a najjednoduchších programov v danom jazyku. Napríklad v jazyku C by program, ktorý na obrazovku vypíše vetu „Hello, world!“, vyzeral nasledovne:

```
#include <stdio.h>
int main(void) {
        printf( "Hello, world!\n" );
        }
```

V oblasti hlbokého učenia by takým jednoduchým príkladom mohol byť multiplayer perceptron. Ide o základný model, ktorý názorne vyjadruje nielen problematiku činnosti a učenia sa neurónových sietí (NN), ale dáva tušiť i komplexnosť, ktorá sa skrýva za oveľa zložitejšími modelmi a technológiami, medzi ktoré patria aj najnovšie generatívne systémy umelej inteligencie.

Príkladom môže byť NN na rozpoznávanie písaných číslic. Sieť má na vstupe rukou napísané číslice v rozlíšení 28x28 pixelov, pričom každý pixel vyjadruje jas daného bodu (body, cez ktoré prechádzalo pero, sú tmavšie, ostatné sú svetlé vo farbe papiera).

Vzhľadom na zameranie tejto publikácie sa pri vysvetľovaní príkladu budeme snažiť o maximálnu jednoduchosť a minimálne ponorenie sa do matematických operácií (i keď matematika stojaca za neurónovými sieťami je síce zaujímavá, ale nie je to žiadna „raketová“ veda:-)[128]

---

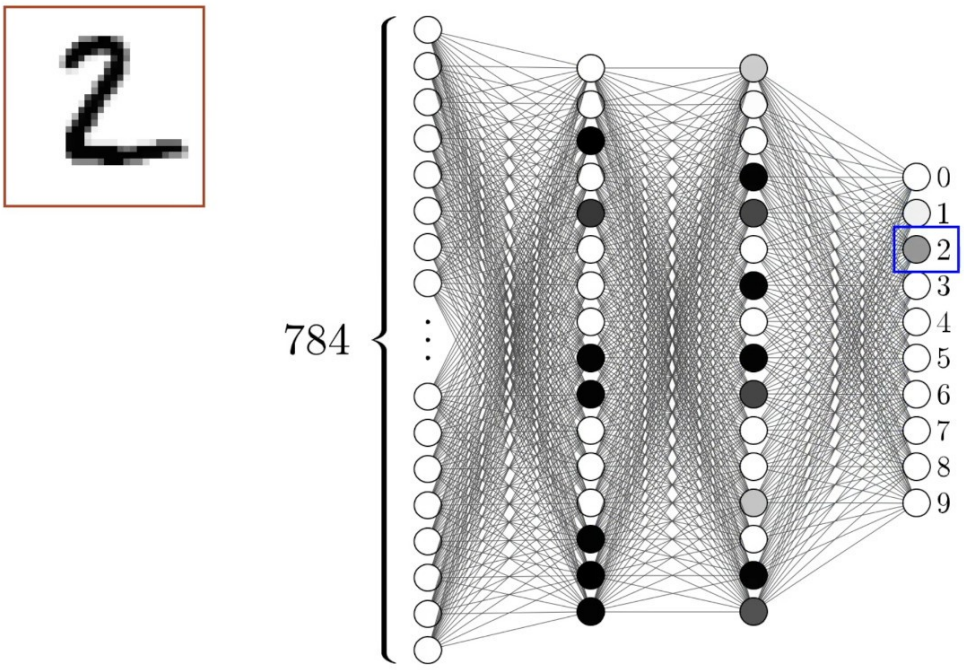
127 **Podobnosť procesov hlbokých neurónových sietí s procesmi prebiehajúcimi v mozgu nás nesmie viesť k zjednodušujúcim záverom**. Problémy a rozdiely uvádzame napr. v kapitole 1.7.

128 Použitý príklad je súčasťou výborného a názorne spracovaného seriálu videí o neurónových sieťach a LLM od 3blue1brown. Prípadným záujemcom o hlbší pohľad a matematickú prezentáciu môžeme tento seriál len odporučiť.

3blue1brown: Neural Networks. [on-line]. [cit. 15. november 2024].

Rozlíšenie na vstupe nám definuje počet vstupných neurónov (28x28=784). K tejto prvej vrstve pridajme dve vnorené vrstvy po 16 neurónoch a výstupnú vrstvu s 10 neurónmi, z ktorých každý vyjadruje pravdepodobnosť, s akou neurónová sieť zistí, že ide o konkrétnu číslicu na vstupe (10 neurónov, t.j. číslice 0-9). Neuróny susediacich vrstiev sú medzi sebou prepojené každý s každým.

Konfiguráciu takto zvolenej neurónovej siete vyjadruje obrázok č. 6.



Obr. č. 6: Príklad NN na rozpoznávanie písaných číslic.[129]

Ako sme uviedli, ide o jeden zo základných príkladov neurónovej siete. I napriek tomu táto sieť obsahuje 13002 parametrov! Rátajte s nami, čo všetko treba zahrnúť:[130]

- váhy medzi vstupnou a 1. vnorenou vrstvou: 784 x 16 = 12544
- váhy medzi 1. a 2. vnorenou vrstvou: 16 x 16 = 256
- váhy medzi 2. vnorenou a výstupnou vrstvou: 16 x 10 = 160
- prahy (bias) neurónov 1. a 2. vnorenej vrstvy i výstupnej vrstvy: 16 + 16 + 10 = 42

Dostupné na internete: <https://youtu.be/aircAruvnKk>

129 Snímka obrazovky, upravené autorom.

3blue1brown: Neural Networks. [on-line]. [cit. 15. november 2024].

Dostupné na internete: <https://youtu.be/aircAruvnKk>

130 V kapitole 1.8 budeme hovoriť o miliardách a biliónoch parametrov generatívnych modelov AI.

Z uvedeného jednoduchého príkladu asi chápeme prečo...

Váhy vyjadrujú silu spojenia neurónu s jednotlivými neurónmi v predchádzajúcej vrstve a prah vyjadruje, do akej miery by mal byť neurón aktívny alebo neaktívny.[131]

Učenie tejto neurónovej siete spočíva v nájdení takých parametrov (váh a prahov), ktoré zabezpečia pre akúkoľvek číslicu na vstupe správnu hodnotu neurónov na výstupe.

Pre prenos signálov medzi jednotlivými vrstvami je dôležitá aktivačná funkcia neurónu. Napríklad aktivačná funkcia pre jeden neurón 1. vnorenej vrstvy môže mať tvar:

$$a_0^{(1)} = \boldsymbol{\sigma} \left(w_{0,0}.a_0^{(0)} + w_{0,1}.a_1^{(0)} + w_{0,2}.a_2^{(0)} + ... + w_{0,783}.a_{783}^{(0)} + b_0\right)$$

kde:

b_0 je prah neurónu 0 z 1. vnorenej vrstvy (neuróny 0-15)

$a_x^{(0)}$ je hodnota aktivačnej funkcie neurónu X z vrstvy 0 (neuróny 0-783)

$w_{0,x}^{(0)}$ je váha medzi neurónom 0 z 1. vnorenej vrstvy a neurónom X z vrstvy 0

$\boldsymbol{\sigma}$ je funkcia sigmoid, resp. iná aktivačná funkcia (napr. ReLU).[132]

Uvedená funkcia ráta len hodnotu jedného neurónu v 1. vstupnej vrstve. Pre všetky neuróny tejto vrstvy používame zápis:

$$\mathbf{a}^{(1)} = \boldsymbol{\sigma} \left(\mathbf{W}\, \mathbf{a}^{(0)} + \mathbf{b} \right)$$

kde:

$\mathbf{a}^{(1)}$ je vektor 16 hodnôt aktivačných funkcií neurónov 1. vnorenej vrstvy

W je matica 784 x 16 hodnôt váhových stavov medzi neurónmi vstupnej a 1. vrstvy

$\mathbf{a}^{(0)}$ je vektor 784 hodnôt aktivačných funkcií neurónov 0. vnorenej vrstvy

b je vektor 16 prahových hodnôt neurónov 1. vnorenej vrstvy

$\boldsymbol{\sigma}$ je aktivačná funkcia vykonaná pre každý jeden neurón 1. vnorenej vrstvy

---

131 Prah (bias) je hodnota, ktorá sa priráta (odráta) k súčtu vážených stavov na zmenu aktivačnej hodnoty neurónu. Prah teda určuje, aká musí byť hodnota váženého stavu, aby sa neurón stal aktívnym. Takto je možné vytvárať napríklad konkrétne vzory, ktoré jednotlivé neuróny detegujú.

132 Definícia sigmoidu: $\boldsymbol{\sigma}(a) = 1 / (1 + e^{-a})$ a definícia ReLU: ReLU (a) = max(0,a)
V moderných NN sa skôr používa ReLU (resp. niektoré iné aktivačné funkcie), nakoľko sigmoid vykazuje problémy v učení, ktoré môžu viesť k nemožnosti v niektorých scenároch natrénovať NN. ReLU funguje výborne osobitne pre extra hlboké neurónové siete.

Neuróny výstupnej vrstvy využívajú ako aktivačnú funkciu tzv. softmax, ktorej výstupom je rozdelenie pravdepodobnosti. Napríklad pre rukou písanú číslicu 2 na vstupe by 2. neurón výstupnej vrstvy mohol mať hodnotu 0,95 a ostatné neuróny medzi 0 a 0,1 (t.j. sieť s 95% pravdepodobnosťou určila, že na vstupe je číslica 2).

Ako sme uviedli, proces učenia je hľadaním správnych parametrov (váh a prahov) našej neurónovej siete. V procese učenia sa využíva veľká množina označených trénovacích dát, ktorých každá položka je tvorená rukou písanou číslicou a správnym označením jej hodnoty. Rukou písané číslice sú v jednotlivých iteráciách zadávané na vstupe NN, pričom na výstupnej vrstve sa pre každý vstup kontroluje, či hodnota výstupných neurónov korešponduje s označením hodnoty danej číslice zo vstupu.

Proces učenia

V procese učenia sa začína s náhodne nastavenými váhami a prahmi, pričom v prvých iteráciách sú výstupné hodnoty náhodné. Preto sa zavádza nákladová funkcia (cost function), ktorej výstup informuje systém o kvalite výsledku danej iterácie. Nákladová funkcia je definovaná ako súčet druhých mocnín rozdielov medzi hodnotami výstupov danej iterácie a hodnotami očakávaného výstupu:

$$C = \mathbf{\Sigma} \left[(a_0^{(k)} - y_0)^2 + (a_1^{(k)} - y_1)^2 + ... + (a_n^{(k)} - y_n)^2 \right]$$

kde:

^{k} je výstupná vrstva NN (napr. softmax) a $_n$ je počet neurónov v tejto vrstve.

a_x predstavuje aktuálny stav konkrétneho neurónu na výstupe

y_x predstavuje očakávaný stav konkrétneho neurónu na výstupe

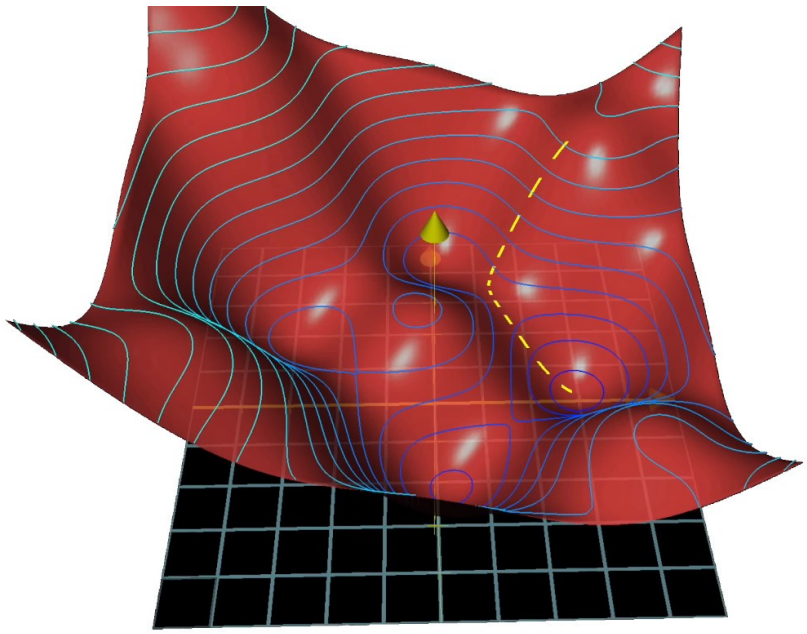
C vyjadruje náklad jedného tréningového príkladu. Čím je náklad nižší, tým je menší rozdiel medzi očakávaným a skutočným výstupom neurónovej siete a tým je sieť presnejšia.

Nákladovú funkciu rátame nielen pre jeden tréningový príklad, ale pre všetky. Priemerný náklad zo všetkých tréningových príkladov je meradlom nedokonalosti a nedostatočného natrénovania neurónovej siete.

Takto definovaná nákladová funkcia je ďalšou vrstvou zložitosti nad neurónovou sieťou, pričom jej výstupom je konkrétne číslo vyjadrujúce kvalitu aktuálneho stavu neurónovej siete (stavu s aktuálnym nastavením váh a prahov pre všetky trénovacie dáta). C je teda

funkciou váh a prahov (existuje medzi nimi funkčná závislosť).

Ak našu úvahu posunieme ďalej, **nákladovou funkciou prevádzame problém hľadania optimálnych váhových stavov a prahov na problém hľadania lokálneho minima funkcie, ktorá má ako vstup všetky váhy a prahy a len jeden výstup (náklad)**. V zásade ide o multidimenzionálnu funkciu, u ktorej hľadáme minimum, t.j. najnižší náklad a tým aj najvyššiu presnosť systému.



Obr. č. 7: Hľadanie lokálneho minima nákladovej funkcie.[133]

Môže ísť o pomerne zložitú funkciu, u ktorej môžu existovať lokálne minimá, z ktorých minimálne jedno nie je problém odhaliť, avšak nájdenie globálneho minima môže byť extrémne náročný problém.

Ide v podstate o hľadanie vektora (má smer a veľkosť), ktorý nás z aktuálneho nákladu čo najefektívnejšie privedie do (aspoň lokálneho) minima. Vektor (smer a veľkosť/strmosť) najstrmejšieho stúpania, resp. v záporných hodnotách klesania, je vyjadrený gradientom danej nákladovej funkcie ($\nabla \mathbf{C}$).

Pri učení sa na základe vyrátaného gradientu $\nabla \mathbf{C}$ vykoná malý krok v smere $-\nabla \mathbf{C}$ a celý proces sa znovu a znovu opakuje.

Dostávame tak algoritmus, pomocou ktorého by sme mohli znížiť náklad, t.j. rozdiel medzi

133 Snímka obrazovky, upravené autorom.

3blue1brown: Neural Networks. [on-line]. [cit. 15. november 2024].

Dostupné na internete: <https://youtu.be/aircAruvnKk>

očakávaným a skutočným výstupom neurónovej siete. Inak povedané, na základe gradientu by sme mohli upravovať váhové stavy a prahy tak, aby bola neurónová sieť presnejšia. Gradient -∇**C** je vektor zmien, ktoré treba prirátať k všetkým váham a prahom systému na dosiahnutie čo najrýchlejšieho poklesu nákladovej funkcie.

A keďže nákladovú funkciu rátame ako priemer nákladov pre všetky trénovacie dáta (cost function of the network), jej minimalizovanie znamená lepšiu presnosť systému pre všetky tréningové príklady.

Algoritmus pre efektívne rátanie gradientu nákladovej funkcie sa nazýva backpropagation algoritmus. Rátanie tohto gradientu patrí k jadru učenia sa nielen neurónových sietí, ale aj viacerých iných metód strojového učenia.

Z doterajšieho opisu činnosti a učenia NN vyplývajú viaceré skutočnosti:

- na jednej strane schopnosť neurónovej siete vykonávať funkcie učenia formou hľadania lokálneho minima, čo je strojovo vykonateľná úloha. Výsledkom je pomerne slušná presnosť činnosti systému pre plánované dáta na vstupe (napr. rukou písané písmená a číslice). Na druhej strane tento princíp činnosti si ťažko poradí so vstupom, ktorý nie je „plánovaný“ (napr. miesto rukou písaného písmena na vstupe je len šum bodov s rôznou farbou a jasom) a v tomto prípade systém nezmyselne chybuje.
- systém, ktorý rozpoznáva a klasifikuje vstupy na základe nákladovej funkcie a hľadania jej minima, nedokáže vykonať opačný proces (t.j. ak je na vstupe číslo 5, tak ho ako päťku klasifikuje, ak však na vstupe zadám príkaz, aby nakreslil číslo 5, tak to nebude vedieť realizovať).

Podľa toho, ako si systém nevie poradiť s „neplánovaným“ vstupom, črtá sa odpoveď na otázku, či systém vôbec chápe, čo klasifikuje, alebo len *memoruje* naučené. Takže je jasné, že neurónová sieť nevie, čo obraz na vstupe znamená...[134]

Určitý pokrok v tejto oblasti badať v technológiách, ktoré sprevádzajú aktuálny rozvoj generatívnych systémov AI, ktorých sémantická a kontextová analýza posúva schopnosti umelej inteligencie bližšie k emulácii myslenia človeka.

Backpropagation algoritmus – algoritmus spätného šírenia

Ako sme uviedli, gradient -∇**C** je vektor zmien, ktoré treba prirátať k všetkým váham

134 Čím viac chápeme, čo jednotlivé skryté vrstvy robia, tým menej a menej inteligencie v tom vidíme:-)

a prahom systému na dosiahnutie čo najrýchlejšieho poklesu nákladovej funkcie.

Vektor $-\nabla \mathbf{C}$ obsahuje jednotlivé hodnoty relatívnej veľkosti, pričom tieto veľkosti vyjadrujú rozdiely medzi dôležitosťou zmien jednotlivých váh a prahov. Zaujímajú nás a pracujeme hlavne s najväčšími hodnotami jednotlivých položiek vektora (teda zmenou váh a prahov dosahujúcich najväčšie hodnoty), lebo tie majú najväčší vplyv na ceste k minimalizovaniu nákladovej funkcie.

Backpropagation je algoritmus, pomocou ktorého na základe rozdielu výstupu pre jeden vstupný príklad rátame požadovaný posun v jednotlivých parametroch modelu. Ide o algoritmus na určenie takých proporčných zmien vo váhach a prahoch, ktoré nás čo najskôr privedú k poklesu nákladovej funkcie.

V zásade môžeme meniť váhy, stavy neurónov predchádzajúcej vrstvy a prahy.

Ak by sme chceli, aby neurónová sieť, ktorá má na vstupe číslicu 2, mala na výstupe $a_2^k=1$ (t.j. neurón výstupnej vrstvy identifikujúci rozpoznanie číslice 2 vykazoval 100%) a hodnoty ostatných neurónov boli nulové, môžeme meniť parametre medzi jednotlivými vrstvami:

- prah pre a_2^k zvýšime, pre ostatné neuróny ^{k} vrstvy znížime.
- pre zmenu hodnoty neurónu a_2^k vo vrstve ^{k} ide o zmenu tých váh medzi a_2^k a neurónmi a_{0-n}^{k-1} predchádzajúcej vrstvy $^{k-1}$, ktoré sa najviac môžu podieľať na zmene hodnoty neurónu a_2^k k požadovanej hodnote (napríklad ak by z neurónov a_{0-n}^{k-1}, mali najväčšiu hodnotu a_2^{k-1}, a_5^{k-1}, a_{10}^{k-1}, menili by sme váhy spojení medzi týmito tromi neurónmi a a_2^k). Váhy teda meníme proporčne podľa hodnoty/stavu jednotlivých neurónov – v zásade len modifikujeme tie, ktoré majú najväčší vplyv na zmenu neurónu a_2^k do požadovaného stavu.
- analogicky môžeme postupovať neurónovou sieťou a meniť stavy neurónov (aktivácie) predchádzajúcich vrstviev podľa jednotlivých váh, t.j. zvyšovať stavy pre kladné hodnoty váh a znižovať pre záporné. A samozrejme toto všetko konáme proporčne podľa veľkosti hodnôt jednotlivých stavov a váh.

Tento proces je treba vykonať pre každý jeden neurón vrstvy ^{k} (t.j. neuróny a_{0-n}^k), pričom výsledné zmeny prahov a váh medzi vrstvami ^{k} a $^{k-1}$ sú priemerom jednotlivých korešpondujúcich hodnôt.

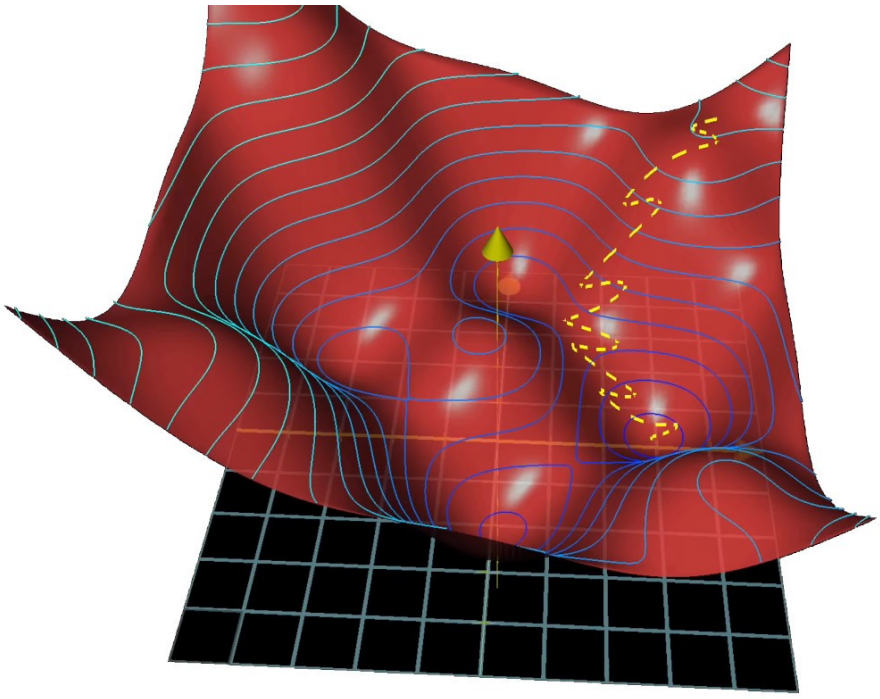
Keďže stavy neurónov a_{0-n}^{k-1} vrstvy $^{k-1}$ nevieme meniť priamo, musíme meniť prahy jednotlivých neurónov a váhy z prepojenia na predchádzajúcu vrstvu $^{k-2}$. Takto postupne

prejdeme celú neurónovú sieť a všetky jej parametre (váhy a prahy, v našom príklade všetkých 13002 parametrov).

Ak by sme tento proces konali len pre číslicu 2 na vstupe, pravdepodobne by sme naučili našu sieť všetky vstupné číslice klasifikovať ako 2:-)

Preto celý tento backpropagation algoritmus aplikujeme na všetky tréningové príklady (backpropagation vyrátame pre každý jeden vstup) a zo všetkých posunov jednotlivých váh vyrátame priemery. Tak dostávame krok klesania (vektor parametrov) gradientu $-\nabla\mathbf{C}$, resp. vzhľadom na to, že pracujeme len s parametrami, ktoré sa najviac podieľajú na požadovanej zmene nákladu, hovoríme o $-\eta\nabla\mathbf{C}$, ktorý je úmerný skutočnej hodnote kroku klesania.

Ak sa nad opísaným algoritmom zamyslíme, uvedomujeme si, že nejde o extra náročné matematické operácie, avšak ide o operácie, ktoré pri reálnych počtoch parametrov neurónových sietí sú extrémne výpočtovo a časovo náročné.



Obr. č. 8: Stochastic gradient descent.[135]

Preto sa tréningové dáta vždy náhodne premiešajú a rozdelia do menších celkov (mini-

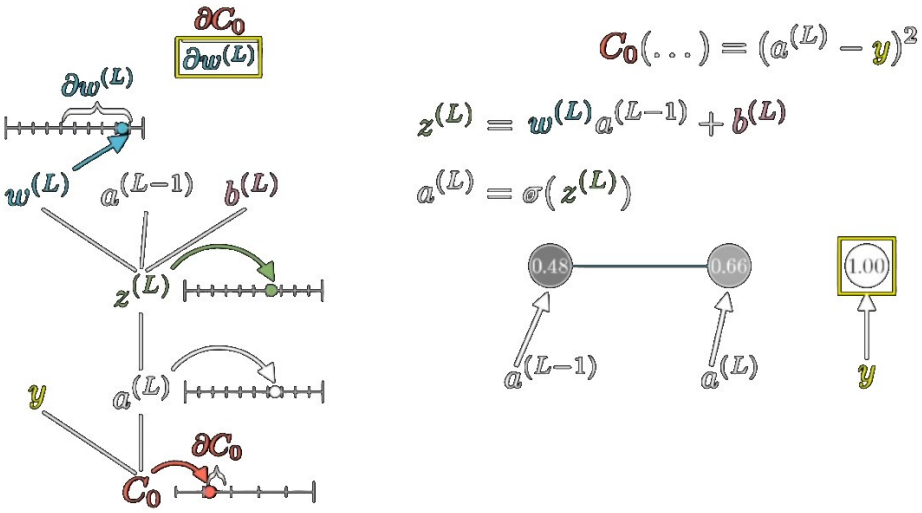
135 Snímka obrazovky, upravené autorom.

3blue1brown: Neural Networks. [on-line]. [cit. 15. november 2024].

Dostupné na internete: <https://youtu.be/aircAruvnKk>

batches, napr. rozdelenie na stovky tréningových vstupných množín), nad ktorými sa samostatne vykonávajú jednotlivé kroky backpropagation algoritmu. Výsledkom je oveľa rýchlejší výpočet, pri ktorom však dochádza k náhodnejšiemu a – geometricky vyjadrené – kľukatejšiemu zostupu k lokálnemu minimu. Ide však o oveľa rýchlejší (cost surface) zostup než priamy, ale veľmi pomalý zostup. Táto metóda sa nazýva *stochastic gradient descent*.

Detailný matematický opis algoritmu spätného šírenia nie je predmetom tohoto diela, preto záverom uvádzame len schematické vyjadrenie základnej matematiky, ktorá je súčasťou iterácie na ceste k požadovaným parametrom navrhnutého modelu neurónovej siete.



Obr. č. 9: Algoritmus spätného šírenia (backpropagation).[136]

Pre čitateľa, ktorý sa bežne nepohybuje v problematike algoritmov umelej inteligencie, sa uvedený príklad môže javiť ako náročný. Ide však len o názornú ukážku postupov, ktoré sa nachádzajú pod vznešenou pokrievkou s názvom hlboké neurónové siete.

V porovnaní s jednoduchosťou programu „Hello, World!“ sa softvérová implementácia i obyčajného multiplayer perceptronu javí ako veľmi zložitý program. I napriek nepomernej zložitosti však systémy AI majú svoj význam, nakoľko riešia problémy, ktoré presahujú schopnosti bežných programov – v duchu hesla „ak je to ťažké, resp. nemožné priamo naprogramovať, nech sa to stroje naučia“.

V súčasnosti je problematika neurónových sietí veľmi prepracovaná, sú riešené nielen

136 Snímka obrazovky, upravené autorom.
3blue1brown: Neural Networks. [on-line]. [cit. 15. november 2024].
Dostupné na internete: <https://youtu.be/aircAruvnKk>

základné problémy učenia (pretrénovanie, strata gradientu a pod.), ale i pokročilé techniky fungovania, dizajnu a trénovania najrozličnejších typov týchto sietí.[137]

1.6. Základné algoritmy strojového učenia

Podľa toho, ako konkrétne proces učenia prebieha, môžeme algoritmy strojového učenia rozdeliť do nasledovných skupín:

- učenie s učiteľom (supervised machine learning)
- učenie bez učiteľa (unsupervised machine learning)
- učenie formou odmeňovania (reinforcement learning)

1.6.1. Učenie s učiteľom (supervised machine learning)

Systémy AI využívajúce učenie s učiteľom dlhý čas tvorili väčšinu systémov strojového učenia. Základom je dostatočne veľká množina správne označených/klasifikovaných trénovacích dát (labeled training dataset) s pozitívnymi a negatívnymi príkladmi. Napr. ak chceme naučiť systém rozpoznávať rôznym spôsobom napísanú číslicu 5, musíme mať veľkú kolekciu napísaných 5-iek, pri ktorých je uvedené označenie (label), že ide o číslicu 5, a tiež veľkú kolekciu iných písaných číslic s označením, že to nie je číslica 5. Tieto kolekcie tvoria tzv. trénovacie dáta (training set/dataset). Systém v procese učenia vyhodnocuje každú jednu predloženú číslicu, pričom výsledok sa porovná s jej označením (je – nie je to 5). V prípade nezhody, sa generuje signál (supervision signal), označujúci nesprávny výsledok, resp. mieru nezhody. Na základe takýchto signálov si systém mení nastavenia váh a prahových hodnôt, až kým sa nedostane do stavu, že všetky vstupné dáta z tréningovej množiny nie sú správne vyhodnotené (v našom prípade všetky číslice 5 vyhodnotené ako 5 a ostatné vyhodnotené ako niečo iné).[138]

137 Prakticky celé portfólio základných algoritmov je pokryté predpripravenými knižnicami vo viacerých softvérových balíkoch (frameworsk) určených na riešenie systémov AI, napr. PyTorch od Meta AI alebo TensorFlow od Google Brain. Podobné je to i so základnými datasetmi (množinami trénovacích dát).

138 Cyklus prijatia informácie na vstupe – jej spracovania – vyhodnotenia výsledku – následná zmena váhových stavov sa pri zložitejších neurónových sieťach nazýva epochou trénovania systému AI. Pri trénovaní systém prechádza mnohými epochami, ktorých ovocím sú meniace sa váhové stavy i parametre systému a zlepšujúca sa klasifikácia, t.j. lepšie výsledky. Nakoniec systém „konverguje“, t.j. medzi jednotlivými epochami sa váhové stavy takmer vôbec nemenia a neurónová sieť je v princípe v rozsahu trénovacích dát vytrénovaná (naučená).
Por. MITCHELL, *Artificial Intelligence*, s. 80.

Ide samozrejme o schematický opis fungovania množiny algoritmov pre supervised machine learning, ktorých presný popis nie je cieľom tohto diela.

1.6.2. Učenie bez učiteľa (unsupervised machine learning)

Učenie bez učiteľa sa používa pri dátach, ktoré neboli vopred klasifikované. Chýba nám teda označenie (label) – správna odpoveď, ktorá klasifikuje vstupné dáta (napr. ide o číslicu 5, na vstupe je trojuholník,...).[139]

Systémy s učením bez učiteľa sa využívajú v prípade, že na vstupe nemáme klasifikované dáta, alebo ide o dáta, ktoré nepoznáme, takže nevieme natrénovať systém na základe dát, ktoré by sme poznali a vedeli ich klasifikovať, resp. označiť. Základné algoritmy využívajúce učenie bez učiteľa teda nie sú schopné určovať správny výstup – ich cieľom je skôr skúmanie a analýza vstupných dát v snahe objaviť a popísať vzory a štruktúry v týchto dátach, teda dozvedieť sa o dátach, resp. z týchto dát niečo viac.

Učenie bez učiteľa sa využíva na riešenie úloh zhlukovania a asociovania.

Cieľom zhlukovania je spájanie dát do skupín, ktoré majú niečo spoločné, napr. segmentácia zákazníkov do skupín s podobnými preferenciami, niektoré problémy spracovania obrazu a detekcie objektov a pod.

Asociovanie sa využíva pri vyhľadávaní asociačných pravidiel, ktoré popisujú množiny dát a vyjadrujú vzťahy medzi nimi. Napr. predajné systémy, ktoré po natrénovaní vedia zákazníkom kupujúcim si určitý produkt ponúknuť relevantné ďalšie produkty (napr. ponuka periférií pri kúpe notebooku), riešenie niektorých problémov pri jazykových prekladoch a pod.

1.6.3. Učenie formou odmeňovania (reinforcement learning)[140]

Ide o osobitnú kategóriu algoritmov strojového učenia, v ktorom sa trénovanie modelu (agenta) realizuje prostredníctvom interakcie s prostredím metódou pokus-omyl. Základom učenia formou odmeňovania (nazývaného aj učenie s posilnením) sú pravidlá, podľa

139 Por. *Algoritmy strojového učenia II.* [on-line]. [cit. 3. januára 2022].
Dostupné na internete: <https://umelainteligencia.sk/algoritmy-strojoveho-ucenia-ii-ucenie-bez-ucitela/>

140 Výborný text o rôznych metódach a najnovšom vývoji v oblasti reinforcement learning vydal Kevin P. Murphy.
MURPHY, P. K. *Reinforcement Learning: An Overview*. [on-line]. [cit. 9. december 2024].
Dostupné na internete: <https://arxiv.org/abs/2412.05265>

ktorých sa agent môže v danom prostredí správať a odmeňovacia funkcia (funkcia užitočnosti, nákladová funkcia), prostredníctvom ktorej agent vie vyhodnotiť, či vykonané rozhodnutie bolo pre neho správne alebo nie.[141]

Učenie agenta teda neprebieha na žiadnych označených, či neoznačených trénovacích dátach, systém v danom prostredí skúša jednotlivé možnosti a učí sa, ktorá možnosť, resp. kombinácia možností je správna a ktorá nie.

Samozrejme ide len o vyjadrenie princípu – jeho realizácie vo forme algoritmov, ako sú napr. Q-learning či Deep Q-learning sú oveľa sofistikovanejšie a využívajú ich napr. šachové systémy, systémy pre pohyb robotických súprav v reálnom prostredí, tretí stupeň trénovania modelov generatívnych systémov AI a pod.

Pre správne vytrénovanie systému učením formou odmeňovania treba zvyčajne extrémne veľa iterácií, preto – ak je to možné – býva trénovanie riešené vo virtuálnych simuláciách daného reálneho (napr. robotického) systému.

1.6.4. Ďalšie spôsoby delenia algoritmov strojového učenia

Doteraz uvedené základné delenie algoritmov strojového učenia podľa toho, ako konkrétne proces učenia prebieha, môžeme doplniť o ďalšie delenia, resp. vyjadriť aj inak.

Napr. medzi **základné typy strojového učenia** patrí: klasifikácia, regresia, zoraďovanie, posilnené učenie, zhlukovanie, detekcia anomálií, odporúčanie a optimalizácia.

Základné algoritmy klasifikácie sú: baseline modely, naivný Bayesov klasifikátor, logistická regresia, Support Vector Machines, rozhodovacie stromy a ensemble metódy.

Medzi z**ákladné regresné algoritmy** rátame: analytické metódy, gradient descent, SVR, regresné stromy.

Základné zhlukovacie algoritmy sú: K-means, hierarchické zhlukovanie, metódy na určenie počtu zhlukov.

Medzi **v súčasnosti najvyužívanejšie pokročilé algoritmy strojového učenia** v produkčnom prostredí patria: transformery, LSTMs (Long Short-Term Memory), CNNs (Convolutional Neural Networks), GBDTs (Gradient Boosted Trees), K-Means Clustering,

141 Por. *Algoritmy strojového učenia III.* [on-line]. [cit. 4. januára 2022]. Dostupné na internete: <https://umelainteligencia.sk/algoritmy-strojoveho-ucenia-iii-ucenie-formou-odmenovania/>

Naive Bayes, Logistic Regression, Reinforcement Learning, K-Nearest Neighbors.[142]

Viaceré z uvedených algoritmov, resp. zadaní je možné realizovať nielen pomocou neurónových sietí, ale – a to častokrát efektívnejšie – aj pomocou klasických metód.[143]

Mnohé moderné technológie umelej inteligencie v sebe kombinujú viacero prístupov – napríklad veľké jazykové modely, medzi ktoré patrí aj ChatGPT, v rámci svojich algoritmov využívajú učenie s učiteľom i bez učiteľa a tiež učenie formou odmeňovania. Umelá inteligencia AlphaGo (prelomová AI, ktorá porazila najlepšieho hráča v hre Go), resp. jej oveľa sofistikovanejší nástupca AlphaZero v sebe kombinujú viaceré metódy, napr. učenie s učiteľom, učenie formou odmeňovania, metódu stromu Monte Carlo a pod.

Toľko aspoň v krátkosti k jednotlivým skupinám algoritmov strojového učenia (detailnejšie rozdelenie je uvedené v prílohách), na ktoré sa teraz môžeme pozrieť trochu z nadhľadu...

1.7. Stačí súčasné strojové učenie, alebo hľadáme ďalej?

Prakticky prvým z algoritmov strojového učenia a ich podmnožiny učenia sa s učiteľom bol Rosenblattov perceptron-learning algorithm. Rosenblatt tiež matematicky dokázal, že pre veľmi špecifickú a limitovanú množinu úloh dokáže správne natrénovaný perceptron (resp. podobný systém učenia sa s učiteľom) vykonávať úlohy bez chýb. Avšak pri iných, resp. všeobecnejších úlohách umelej inteligencie je úspešnosť takéhoto systému diskutabilná.[144] [145] [146]

---

142 REDDY, B. *The Most Popular Machine Learning Algorithms in Production* [on-line]. Twitter. [cit. 6. septembra 2023].
Dostupné na internete: <https://twitter.com/bindureddy/status/1698919098489733212>

143 Napríklad neurónové siete nie sú pre vhodné pre tabulárne dáta (pri klasifikačných zadaniach je vhodnejší napr. Gaussov naivný bayes klasifikátor).

144 Por. MITCHELL, *Artificial Intelligence*, s. 30.

145 MINSKY, M. L., PAPERT, S. L. *Perceptrons: An Introduction to Computational Geometry*. Cambridge, Mass: MIT Press, 1969.

146 Zložitejšie úlohy samozrejme vyžadujú použitie viacvrstvových neurónových sietí, pre ktoré však v tom čase neexistoval učiaci sa algoritmus, t.j. spôsob, ako vo viacerých vrstvách realizovať spätnú väzbu a nastavovanie váhových stavov v procese učenia. Preto Minsky dokonca zavrhol celý koncept subsymbolickej AI. Až neskorší vývoj všeobecného algoritmu učenia sa, tzv. **back-propagation** algorithm (algoritmus spätného šírenia chyby), ktorý v roku 1986 predstavili Rumelhart, Hinton a Williams,

Už pri tak jednoduchej realizácii algoritmu strojového učenia je možné si uvedomiť viaceré riziká spojené so systémami umelej inteligencie, napríklad:[147]

- nesprávna zmena váh a prahových hodnôt v procese učenia (napr. priveľké zmeny váhových hodnôt medzi jednotlivými krokmi učenia) môže viesť k nesprávnemu natrénovaniu systému (systém bude dávať nesprávne výsledky v prevádzke).
- u niektorých skupín učenia je pre realizáciu konkrétnych modelov treba vykonať nesmierne veľa iterácií v procese učenia. Obmedzenie možností učenia vedie k modelom, ktoré dávajú nesprávne, resp. v určitých situáciách nesprávne výsledky.
- obmedzenie na množinu úloh, ktoré dokáže takýto systém riešiť.
- praktická nemožnosť previesť naučené hodnoty váh a prahových hodnôt do ľudsky pochopiteľných pravidiel, systém sa tak stáva „čiernou skrinkou“.

Čim zložitejší systém AI (v súčasnosti dokonca so stovkami miliárd až desiatkami biliárd váhových stavov a prahových hodnôt, s veľkým množstvom vrstiev a parciálnych architektúr neurónových sietí, atď.), tým sú uvedené problémy vo väčšine scenárov vypuklejšie. Vyvstáva tak dôležitá otázka – **sme potom schopní overovať a nastavovať etické pravidlá, resp. všeobecne obmedzenia a mantinely systémov umelej inteligencie?**[148]

Mnohí obhajcovia subsymbolických systémov umelej inteligencie – argumentujúc „architektúrou“ mozgu na úrovni neurónov – sú presvedčení, že prostriedkami symbolickej AI (symboly a logické vzťahy medzi nimi) nie je možné dosiahnuť umelú inteligenciu, keďže symboly v mozgu vznikajú na základe neurálnych procesov (na úrovni neurónov).[149]

I napriek všetkým úspechom, ktoré sme v súčasnosti vďaka subsymbolickým systémom, medzi ktoré patria neurónové siete, systémy hlbokého učenia atď., dosiahli, **dovolíme si**

---

znamenal skutočný prelom vo vývoji neurónových sietí, pričom v spojení s hlbokým učením sa tieto stali základom pre aktuálny veľký pokrok v oblasti AI.

NIELSEN, M. *Neural Networks and Deep Learning*. [on-line]. [cit. 19. januára 2022]. Dostupné na internete: <http://neuralnetworksanddeeplearning.com/>

147 Komplexnej analýze limitov a rizík súčasných systémov umelej inteligencie sa venujeme v 2. kapitole.

148 Nie je to problém pri systéme na rozpoznávanie číslic, ale môže to byť skutočný problém pri automatických bojových systémoch, autopilotov vozidiel, prostriedkov na podporu života, systémov na detekciu teroristických hrozieb, podporných systémov pre súdnictvo, atď.

149 Por. MITCHELL, *Artificial Intelligence*, s. 31.

s názorom, že skutočnú umelú inteligenciu je možné dosiahnuť výlučne prostriedkami subsymbolickej AI, nesúhlasiť, keďže stále nie sme schopní prekročiť hranice ANI, t.j. hranicu medzi úzko špecializovanými systémami (slabej) umelej inteligencie a všeobecnými systémami AGI.[150] Radi by sme uviedli aspoň dve výhrady...

Domnievame sa, že **ak budujeme systémy umelej inteligencie inšpirujúc sa ľudským mozgom, resp. sa chceme priblížiť analógii s mozgom, je treba vziať do úvahy metódy subsymbolickej i symbolickej AI**. Na jednej strane by tak išlo o postupy, ktoré sú jednou z možností ďalšieho rozvoja umelej inteligencie,[151] na druhej strane – čo je pre nás podstatné – by išlo o systémy, v ktorých na úrovni symbolickej AI by sme azda boli schopní ľahšie riešiť etické výzvy a implementovať pravidlá aj v prípade komplexných a pokročilých systémov AI.[152]

Aj keď – ako sme uviedli v kapitole 1.4. – logika prvého rádu sa nedokázala vysporiadať s neistotou a nedeterministickým vnímaním sveta, nesmieme ju vylúčiť z dizajnu systémov AI. V súčasnosti sa totiž robí opačný extrém – na základe preferovania subsymbolických systémov sa rezignovalo na logiku až do tej miery, že moderní vývojári umelej inteligencie o nej mnohokrát ani nechyrujú.[153] **Keďže sa však reálny svet skladá z objektov a vzťahov medzi nimi, logika prvého rádu patrí k fundamentom pri popise reálneho sveta a ako taká by mala byť súčasťou návrhu sofistikovaných systémov AI**.[154] [155]

150 Informácie o ANI a AGI sú uvedené v kapitole 1.3.

151 V súčasnosti sa pokroky v niektorých oblastiach moderných systémov AI dosahujú (niekedy až radikálnym) vylepšovaním existujúcich algoritmov a súčasne kombináciou viacerých systémov, metód a postupov na dosiahnutie cieľa. Napr. v súčasnosti sa v oblasti robotiky integráciou generatívnych systémov AI dosahujú veľké pokroky v ich učení a schopnostiach.

152 Treba však uviesť, že až donedávna pokusy vytvoriť hybridný systém integrujúci subsymbolické i symbolické metódy AI nemali nejaký väčší úspech.
MITCHELL, *Artificial Intelligence*, s. 41.
Určitý pokrok nastáva až s aktuálnym rozvojom generatívnych systémov AI, osobitne v oblasti komplexného uvažovania a niektorých ďalších noviniek genAI, ktoré diskutujeme v kapitole 1.8.2.

153 Pre symbolické systémy postavené na logike sa dokonca zaužíval pejoratívny názov GOFAI – Good Old-Fashioned AI.
HAUGELAND J. *Artificial Intelligence: The Very Idea*. MIT Press, 1985.

154 Por. RUSSELL, *Human Compatible*, s. 271.

155 Rovnaký pohľad zdieľa i Demis Hassabis, zakladateľ a CEO spoločnosti DeepMind venujúcej sa najpokročilejším systémom AI: „**Na dosiahnutie skutočnej inteligencie nestačia súčasné systémy hlbokého učenia, ktoré sú ekvivalentami zmyslových centier v mozgu, napr. vizuálneho alebo**

Na základe doterajšieho výskumu v oblasti AI môžeme povedať, že systémy schopné zmysluplne poznávať veci musia mať kapacitu reprezentovať ich a uvažovať nad nimi porovnateľnú minimálne s logikou prvého rádu, pričom zatiaľ nevieme, akú presnú podobu budú tieto systémy mať. Ich logika totiž môže byť začlenená do pravdepodobnostných systémov uvažovania, do systémov hlbokého učenia alebo do nejakého hybridného dizajnu, ktorý ešte len bude vynájdený.[156]

Druhou výhradou je fakt, že i napriek veľkej snahe dnešné hlboké neurónové siete a algoritmy hlbokého učenia (teda dnešné subsymbolické systémy) len čiastočne simulujú mozgovú činnosť a prácu neurónov. **Biologické neuróny sa správajú inak ako uzly umelých neurónových sietí a mozgové procesy sú v mnohom komplexnejšie**.

Nepochybne medzi biologickými a umelými neurónmi i celými neurónovými sieťami existuje veľká podobnosť:[157]

biologický (reálny) neurón	**umelý neurón (uzol)**
synaptická účinnosť	váha spojenia (w_i)
séria akčných potenciálov	výstupná aktivita (y)
postsynaptický potenciál	$w_i * x_i$
sumárny synaptický potenciál	$\mathbf{\Sigma}$ ($w_i * x_i$)
prah excitácie	$\boldsymbol{\sigma}$
synaptická plasticita	Δw_i (zmena synaptickej váhy)

Avšak **rozdielov je viac, než by sme si želali**...

Uzol napodobňujúc činnosť neurónu vysiela impulz, resp. vo všeobecnosti číselné stavy.

---

sluchového. Skutočná inteligencia je však oveľa viac než to, keďže je treba skombinovať zmyslové vstupy do myslenia vyššej úrovne a symbolického uvažovania. Musíme preto tieto systémy budovať až do symbolickej úrovne uvažovania a myslenia - t. j. do úrovne matematiky, jazyka a logiky.“

HEATH N. *Google DeepMind founder Demis Hassabis: Three truths about AI* [on-line]. TechRepublic, September 24, 2018. [cit. 14. augusta 2022].

Dostupné na internete: <https://www.techrepublic.com/article/google-deepmind-founder-demis-hassabis-three-truths-about-ai/>

156 RUSSELL, *Human Compatible*, s. 272.

157 JEDLIČKA, P. et al. *Molekulové mechanizmy učenia a pamäti*. In: Hulín, I. (ed.). *Patofyziológia* [6. vyd.]. [on-line]. Bratislava: Slovak Academic Press, 2002, s. 1183-1199. [cit. 24. júla 2025].

Dostupné na internete: <http://ii.fmph.uniba.sk/~benus/books/MolMechanizmy.pdf>

Výstupom biologického neurónu je však elektrický prúd spôsobený pohybom iónov v bunke. Jednotlivé výstupné prúdy sa prostredníctvom synáps posúvajú do ďalších buniek, kde zvyšujú ich membránový potenciál, ktorý je výsledkom nerovnovážnej koncentrácie iónov. Keď membránový potenciál neurónu prekročí určitú prahovú hodnotu, bunka vyšle impulz (akčný potenciál / „spike“), t. j. prúd, ktorý sa má odovzdať do ďalších buniek.[158] V biologických neurónoch okrem akčných potenciálov nachádzame aj iné signály a kontinuálnu aktivitu (napr. synaptické potenciály a pod.). V umelých neurónových sieťach je tento proces simulovaný len výstupným binárnym impulzom a váhovým stavom „synapsy“ spájajúcej dva uzly.

Biologické neuróny majú naviac vnútornú dynamiku, ktorá spôsobuje ich zmeny v čase. S plynutím času majú tendenciu vybíjať sa a znižovať svoj membránový potenciál, čoho dôsledkom je napríklad skutočnosť, že riedke impulzy na vstupe neurónu nespôsobia výstupný impulz![159] Okrem tejto nelineárnej dynamiky u biologických systémov pozorujeme i tzv. oscilácie, t.j. vlny neuronálnej aktivity, a tiež distribuované a paralelné aktivácie rôznych mozgových centier.[160]

Dendrity v mozgu nefigurujú len ako „spojnice“ určitej sily medzi jednotlivými neurónmi, ale obsahujú zložité mechanizmy založené na iónových kanáloch, ktoré samy o sebe vedia robiť výpočty – okrem lineárnych (napr. 1+1=2) v nich prebiehajú aj nelinearity, takže vstup môže byť mnohorako spracovaný. Už v samotných dendritoch môžu byť napríklad implementované logické funkcie, ako „a/and“ či „alebo/or“.[161]

Odlišná je i synaptická plasticita mozgových neurónov v porovnaní s dynamikou zmeny váhových stavov uzlov umelých neurónových sietí.

---

158 *Action potential* [on-line]. [cit. 27. februára 2022].
Dostupné na internete: <https://en.wikipedia.org/wiki/Action_potential>

159 KORAKOVOUNIS, D. *Spiking Neural Networks: where neuroscience meets artificial intelligence*. [on-line]. [cit. 27. februára 2022].
Dostupné na internete: <https://theaisummer.com/spiking-neural-networks/>

160 Por. JEDLIČKA et al., *Molekulové mechanizmy učenia a pamäti*, s. 1194-1196.

161 Por. JEDLIČKA et al., *Molekulové mechanizmy učenia a pamäti*, s. 1187.
LIPTÁK, J. *Neurovedec Peter Jedlička: Mozgová bunka dokáže oveľa viac ako počítačový neurón. Nekontrolovanej umelej inteligencie sa neobávam.* [on-line]. [cit. 24. júla 2025].
Dostupné na internete: <https://www.postoj.sk/177474/mozgova-bunka-dokaze-ovela-viac-ako-pocitacovy-neuron-nekontrolovanej-umelej-inteligencie-sa-neobavam>

Dôležitým rozdielom je i spôsob komunikácie medzi biologickými neurónmi a medzi uzlami. Biologické neuróny komunikujú a spracúvajú informácie asynchrónne, kým typické uzly neurónových sietí synchrónne, t.j. v jednom kroku všetky uzly jednej vrstvy neurónovej siete prečítajú vstup, vyrátajú výstup a posúvajú ho ďalej. U biologických neurónov je to iné – v každom okamihu môžu prijať vstupný signál a vytvoriť výstup bez ohľadu na správanie sa ostatných neurónov.[162]

Tiež treba poznamenať, že biologické systémy majú oveľa prepracovanejší systém spätnej väzby na úrovni neurónov (čo je súčasťou procesov učenia sa a adaptívnej činnosti), než sú dnešné metódy používané v neurónových sieťach (napr. uvádzaný algoritmus spätného šírenia chyby).

Biologické neurónové siete sú zvyčajne oveľa viac nelineárne „výpočtové jednotky“ než umelé siete.

Do konkrétnej činnosti sa dokážu zapájať neuróny rôznych mozgových centier, pričom jednotlivé neuróny dokážu plniť rôzne úlohy pre rôzne činnosti. Ukladanie informácií sa môže vykonávať naprieč rôznymi regiónmi v mozgu, pričom jednotlivé neuróny zohrávajú úlohu pri obrovskom množstve rôznych spomienok a pri predstavovaní si budúcnosti.[163]

Rozdielny je i spôsob ukladania a vybavovania si spomienok. Ľudská pamäť uprednostňuje užitočné informácie, ktoré nám dovoľujú fungovať vo svete.[164] Terajšie generatívne systémy AI, ktorých schopnosti sú v súčasnosti najviac porovnávané s inteligenciou človeka, obsahujú niečo na spôsob komprimovaného obrazu sveta a jeho informácií,[165] pričom pri ďalších modifikáciách a dotrénovaní špecifických schopností majú tendenciu niektoré podstatné znalosti strácať.

A aby toho nebolo málo, neuróny v rámci biologických systémov ako každé iné bunky

---

162 KORAKOVOUNIS, *Spiking Neural Networks: where neuroscience meets artificial intelligence*. [on-line]. [cit. 27. februára 2022].
Dostupné na internete: <https://theaisummer.com/spiking-neural-networks/>

163 JOSSELYN, S.A., TONEGAWA, S. *Memory engrams: Recalling the past and imagining the future*. In: Science. [on-line]. 2020, 367(6473), eaaw4325. [cit. 24. júla 2025].
Dostupné na internete: <https://doi.org/10.1126/science.aaw4325>

164 MCLAIN, R. *Can your brain run out of memory?* [on-line]. [cit. 24. júla 2025].
Dostupné na internete: <https://www.livescience.com/health/neuroscience/can-your-brain-run-out-of-memory>

165 Komprimovaný, resp. „zazipovaný“ obraz sveta diskutujeme v závere kapitoly 1.8.2.

majú jadro a v ňom DNA, ktorá s činnosťou týchto neurónov úzko súvisí. Interakcia medzi génmi a ich prostredím ovplyvňuje činnosť neurónov až do tej miery, že sa to môže prejaviť na prenose signálov v neurónových sieťach a v konečnom dôsledku aj na kognitívnych funkciách.[166]

Ak sa pozrieme na energetickú náročnosť, na jednej strane **máme súčasné sofistikované neurónové siete, ktoré majú rastúce, ba až neúnosné výpočtové náklady** (a tým aj vysokú spotrebu energie, časovú náročnosť, a pod.) potrebné na ich kvalitné vytrénovanie a rozvoj.[167] Tieto siete vedia byť veľmi efektívne v konkrétnych scenároch použitia, no ak by mali vedieť riešiť pre človeka jednoduché základné úlohy[168], resp. doterajšie zadania zovšeobecňovať a nebodaj aj chápať, sú v koncoch.

Na strane druhej **máme prirodzenú inteligenciu s nepatrnou spotrebou energie** (len cca. 20 wattov!), schopnú kreativity, riešenia problémov a multitaskingu. Ide o biologické systémy, ktoré si prirodzenou evolúciou osvojili spracovanie informácií a reagovanie na ne.[169] [170]

Uvedené porovnávanie napovedá, že vývoj systémov umelej inteligencie a výskum

166 Siete interakcií medzi génmi a ich prostredím, ktoré ovplyvňujú činnosť bunky, sú vyjadrené génovými regulačnými sieťami (GRN). V rámci bioinformatiky hovoríme o spracúvaní informácií o génoch a proteínoch v biologickej bunke s dôrazom na ich dynamickú interakciu. V bunke a osobitne v neuróne tak dochádza k neustálej interakcii DNA, RNA a proteínov, ktorá ovplyvňuje fungovanie celej bunky a fenotyp organizmu s dôsledkami na výstupné signály a funkcie bunky, resp. neurónu.
Uvedenej problematike sa venuje Computational Neurogenetic Modeling (CNM), kombinujúc počítačovú neurovedu, genetické algoritmy a metódy strojového učenia v snahe lepšie pochopiť komplexnú interakciu medzi génmi a aktivitou neurónov a následne vysvetliť, ako môžu genetické odchýlky ovplyvniť aktivitu neurónov a v konečnom dôsledku aj poznanie a správanie.
BEŇUŠKOVÁ, Ľ., KASABOV, N. *Computational Neurogenetic Modeling*. New York: Springer-Verlag US, 2010, s. 137-153.

167 „The cost of improvement is becoming unsustainable.“
THOMPSON, N. C., GREENEWALD, K., LEE, K., MANSO, G. F. *Deep learning computational cost*. [on-line]. [cit. 27. februára 2022].
Dostupné na internete: <https://spectrum.ieee.org/deep-learning-computational-cost>

168 „The easy things are hard“ – táto problematika je rozoberaná v kapitole 1.10.

169 KORAKOVOUNIS, *Spiking Neural Networks: where neuroscience meets artificial intelligence*. [on-line]. [cit. 27. februára 2022].
Dostupné na internete: <https://theaisummer.com/spiking-neural-networks/>

170 V našom kontexte sú tieto systémy i súčasťou etického rámca, ktorý biológiu presahuje.

v oblasti neurovied (výpočtová neuroveda, neurobiológia,...) majú blízko k sebe. Podobne ako v AI existujú konkrétne algoritmy a modely, v neurovedách sa využívajú tzv. mechanistické modely biologických neurónov a mozgových neurónových sietí.[171] A tak, ako systémy AI pomáhajú v modelovaní a ladení mechanistických modelov biologických neurónov, mechanistické modely inšpirujú vývoj nových návrhov umelých neurónových sietí.[172]

Ovocie pokračujúceho skúmania mozgových procesov a neurónov, snahy pochopiť, čo ich robí tak efektívnymi a odvaha aplikovať tieto zistenia do oblasti umelej inteligencie viedli k vzniku tretej generácie neurónových sietí, ktorú asi najviac reprezentujú tzv. **spiking neural networks** (SNNs) a ich implementácie v rámci neuromorfnej hardvérovej a softvérovej architektúry.[173] [174]

SNNs v súčasnosti predstavujú asi najpokročilejšie **modely neuromorfných sietí**, od ktorých sú odvodené, resp. okrem nich poznáme viaceré ďalšie modely, ako napr.

---

171 V rámci mechanistických modelov je možné veľmi dobre simulovať anatomické, fyziologické, biofyzikálne a biochemické mechanizmy biologickej nervovej bunky pomocou jednoduchého RC obvodu, ktorý bežne poznáme z oblasti elektrotechniky.
HERZ, A.V., GOLLISCH, T., MACHENS, C.K., JAEGER, D. *Modeling single-neuron dynamics and computations: a balance of detail and abstraction*. In: Science. [on-line]. 2006, 314(5796):80-5. [cit. 25. júla 2025].
Dostupné na internete: <https://doi.org/10.1126/science.1127240>
STERRATT, D., GRAHAM, B. et al. *Principles of Computational Modeling in Neuroscience*. Cambridge University Press, 2023. ISBN 978-1-108-71642-0.

172 LINDSAY, G. *Models of the Mind: How Physics, Engineering and Mathematics Have Shaped Our Understanding of the Brain*. Bloomsbury Sigma, 2021. ISBN 978-1-472-96642-1.

173 PFEIFFER, M., PFEIL, T. *Deep Learning With Spiking Neurons: Opportunities and Challenges*. [on-line]. [cit. 27. februára 2022].
Dostupné na internete: <https://www.frontiersin.org/articles/10.3389/fnins.2018.00774/full>

174 Predchádzajúce generácie neurónových sietí sú založené na McCullochových Pittsových neurónoch (t.j. prahových hradlách), resp. sigmoidálnych hradlách, pričom sa javí, že jeden spike neurón dokáže riešiť takú konkrétnu biologickú funkcionalitu, ktorá by si vyžadovala stovky skrytých jednotiek sigmoidálnej neurónovej siete.
MAASS, W. *Networks of spiking neurons: The third generation of neural network models*. [on-line]. [cit. 28. februára 2022].
Dostupné na internete: <https://www.sciencedirect.com/science/article/abs/pii/S0893608097000117>

Pulse-coupled neural networks (PCNNs)[175] pre vysoko efektívne spracovávanie obrazu, Liquid State Machines (LSMs)[176] s adaptabilnou schopnosťou flexibilne a dynamicky sa prispôsobiť meniacim sa vstupom, a pod.

Z najnovších trendov môžeme ešte spomenúť novátorský koncept umelých neurónových sietí využívajúci poznatky z oscilačnej dynamiky v neokortikálnych obvodoch, ktorý by mal smerovať k návrhu energeticky oveľa efektívnejších a úspornejších systémov AI.[177]

Neuromorfné siete, resp. neuromorfné systémy vo všeobecnosti, sú navrhnuté tak, aby napodobňovali štruktúru a fungovanie ľudského mozgu. (resp. vo všeobecnosti mozgu cicavcov). V snahe o **pokročilejšie napodobnenie biologických systémov** majú potenciál vytvoriť efektívnejšie a adaptívnejšie systémy umelej inteligencie.[178]

I napriek tomu, že v súčasnosti prakticky nejestvujú systémy reálneho nasadenia neuromorfných sietí, ide o jednu z perspektívnych technológií, do ktorej sa vkladajú veľké nádeje. Na rozdiel od tradičných neurónových sietí by neuromorfné siete mohli spracúvať informácie v reálnom čase, podobne ako to robí ľudský mozog. Mali by byť schopné učiť sa a adaptovať sa na nové podmienky bez potreby opätovného trénovania.[179] Dôležitým vkladom pre rozvoj pokročilých systémov AI by mala byť aj predpokladaná energetická efektívnosť neuromorfných technológií. Súčasné implementácie sofistikovaných neurónových sietí sú extrémne energeticky a výpočtovo náročné.[180] Neuromorfné siete by

175 ZHAN, K., SHI, J., WANG, H. et al. *Computational Mechanisms of Pulse-Coupled Neural Networks: A Comprehensive Review.* In: *Arch Computat Methods Eng*. [on-line]. 2017, 24, s. 573–588. [cit. 8. júla 2023].
Dostupné na internete: <https://doi.org/10.1007/s11831-016-9182-3>

176 GBENDA, O.G. *Research on the Concept of Liquid State Machine*. [on-line]. Heriot-Watt University, 2019. [cit. 8. júla 2023].
Dostupné na internete: <https://doi.org/10.48550/arXiv.1910.03354>

177 EFFENBERGER, F., CARVALHO, P. et al. *The functional role of oscillatory dynamics in neocortical circuits: A computational perspective*. [on-line]. PNAS, 2025. [cit. 25. júla 2025].
Dostupné na internete: <https://www.pnas.org/doi/pdf/10.1073/pnas.2412830122>

178 *Neuromorphic engineering.* [on-line]. [cit. 8. júla 2023].
Dostupné na internete: <https://en.wikipedia.org/wiki/Neuromorphic_engineering>

179 Ide o významný rozdiel oproti tradičným neurónovým sieťam, ktoré vyžadujú opätovné trénovanie, resp. dotrénovanie pre schopnosť adaptovať sa na zmenené podmienky.

180 Energetickú a výpočtovú náročnosť súčasných systémov AI rozoberáme v kapitole 2.4.

mali byť v tomto smere oveľa efektívnejšie a úspornejšie.[181]

Neuromorfné systémy v súčasnosti predstavujú sľubnú technológiu pre ďalší rozvoj pokročilých algoritmov umelej inteligencie. Na ceste k reálnemu nasadeniu a technologickej dospelosti však musia prekonať viaceré špecifické detské choroby a zdolať viaceré výzvy praktických hardvérových a softvérových implementácií.[182]

Vrátiac sa k začiatku tejto diskusie, t.j. rozporovaniu tvrdenia, že súčasné algoritmy strojového učenia sú schopné dosiahnuť skutočnú umelú inteligenciu, vidíme, že je pred nami ešte dlhá cesta a vývoj, ktorý v kontexte doterajších subsymbolických i symbolických prístupov môže priniesť veľa zaujímavých poznatkov a progresu na ceste (nielen) k silnej a všeobecnej umelej inteligencii.

---

181 SCHUMAN, C.D., KULKARNI, S.R., PARSA, M. et al. *Opportunities for neuromorphic computing algorithms and applications*. In: *Nat Comput Sci*. [on-line]. 2022, 2, s. 10–19. [cit. 8. júla 2023]. Dostupné na internete: <https://doi.org/10.1038/s43588-021-00184-y>

182 V rámci výskumu a hľadania vhodných ciest **hardvérovej implementácie neuromorfných architektúr** sú vyvíjané nielen neuromorfné čipy (napr. Darwin III), ale i celé výpočtové platformy až po úroveň superpočítačov (napr. čínsky Wukong / Darwin Monkey, ktorý s 960 čipmi Darwin III bol v septembri 2025 najvýkonnejším neuromorfným superpočítačom napodobňujúcim architektúru mozgu s viac ako dvomi miliardami umelých neurónov a sto miliardami umelých synapsií, ktoré simulujú neurálnu štruktúru mozgu makaka).
HUDGES, O. *China's 'Darwin Monkey' is the world's largest brain-inspired supercomputer*. [on-line]. [cit. 7. septembra 2025].
Dostupné na internete: <https://www.livescience.com/technology/computing/chinas-darwin-monkey-is-the-worlds-largest-brain-inspired-supercomputer>
Analogickým **softvérovým riešením neuromorfnej architektúry a spiking neural networks** (SNN) je v septembri 2025 predstavený čínsky jazykový model SpikingBrain 1.0. Neuromorfnou emuláciou činnosti mozgu by mal dosahovať rádovo vyššie výkony než terajšie najvýkonnejšie systémy genAI pri súčasnom minimálnom nároku na trénovanie (o dva rády) a oveľa nižšej energetickej náročnosti.
PAN, Y., FENG, Y., ZHUANG, J. et al. *SpikingBrain Technical Report: Spiking Brain-inspired Large Models*. [on-line]. [cit. 15. septembra 2025].
Dostupné na internete: <https://arxiv.org/html/2509.05276v1>

1.8. Generatívne systémy umelej inteligencie (genAI)

Súčasťou druhého, rozšíreného vydania tejto publikácie je i snaha reflektovať vývoj, ktorý nastal po predstavení najnovších generatívnych systémov umelej inteligencie, pričom ich neustále narastajúci technologický a spoločenský dopad a dôsledky si v čase prvého vydania takmer nikto nevedel predstaviť.

30. novembra 2022 spoločnosť OpenAI predstavila ChatGPT – veľký jazykový model, ktorý na základe textových vstupov generoval požadované odpovede.[183] Technologicky vychádzal z 3. generácie modelu GPT (Generative Pre-trained Transformer), ktorý bol predstaviteľom sľubne sa rozvíjajúcej tzv. generatívnej umelej inteligencie. Ako už názov napovedá, **algoritmy generatívnych systémov AI sú schopné generovať rôznorodý obsah** – text, obrázky, audio a video a pod.

ChatGPT (GPT 3.5) spočiatku neznamenal len určitý technologický prielom, ako skôr sociologický a psychologický zlom v prístupe verejnosti k systémom a možnostiam umelej inteligencie. Nastal veľký dopyt po generatívnych systémoch AI a možnosti ich nasadenia i spôsoby využitia presiahli akékoľvek očakávania.[184]

Jadrom veľkých jazykových modelov sú tzv. GPT moduly, ktoré fungujú na základe analýzy vstupnej sekvencie a použitia zložitej matematiky na predpovedanie najpravdepodobnejšieho výstupu. Na určenie najlepšieho možného ďalšieho slova vo vete na základe všetkých predchádzajúcich slov sa využíva pravdepodobnosť a strojová analógia syntaktickej a sémantickej analýzy.[185] GPT modely, ako aplikácia

183 *Introducing ChatGPT.* [on-line]. [cit. 10. januára 2024].
Dostupné na internete: <https://openai.com/index/chatgpt/>

184 Jazykové modely pre spracovanie prirodzeného jazyka (NLP) sú s nami už dlhší čas. Jedným z prvých modelov NLP bol systém Eliza z roku 1996. Na rozdiel od genAI mal predprogramované odpovede založené na kľúčových slovách a porovnávaní vzorov. Mal veľmi obmedzené chápanie anglického jazyka a podobne ako v prípade mnohých prvých jazykových modelov bolo v jeho logike badať evidentné chyby. Od roku 1972 sa súbežne vyvíjali rekurentné neurónové siete (RNN), ktoré tvorili prvú skutočnú technológiu schopnú predikovať nasledovné slovo vo vete. No až s vývojom architektúry transformerov s vlastnosťami ako self attention (2017/2018) prichádza skutočný rozvoj generatívnych systémov AI. BERMAN, M. *Large Language Models (LLMs) - Everything You NEED To Know.* [on-line]. [cit. 23. septembra 2024].
Dostupné na internete: <https://www.youtube.com/watch?v=osKyvYJ3PRM>

185 Na základe štatistickej analýzy obrovského množstva textov sa modely učia syntaktické vzory a vzťahy,

technológie hlbokého učenia umelej inteligencie, **dokážu spracovať výzvy v prirodzenom jazyku a generovať relevantné textové odpovede podobné ľudským**.[186]

Podľa textového vstupu (promptu) GPT vytvorí najpravdepodobnejšiu odpoveď na základe svojich trénovacích dát, ktoré obsahujú miliardy verejne dostupných zdrojov textových údajov od slávnych literárnych diel, verejne dostupných dát z internetu až po otvorený zdrojový kód. **Rozsiahle súbory trénovacích dát umožňujú GPT napodobniť schopnosť porozumieť ľudskému jazyku**. Veľké modely GPT používajú hlboké učenie na spracovanie kontextu a čerpajú znalosti z relevantných textov vo svojich trénovacích údajoch na predpovedanie optimálnej odpovede.[187]

Schopnosti modelov GPT pramenia z dvoch kľúčových aspektov:

- **generatívneho predtrénovania**, ktoré učí model rozpoznávať vzory v neoznačených údajoch a následne tieto vzory aplikovať na nové vstupy;
- **architektúry transformerov**, ktorá umožňuje modelu paralelne spracovať všetky časti vstupnej sekvencie a generovať výstupy.

Generatívne systémy, či už ide o modely jazykové, grafické alebo generujúce video, vedia byť úžasné v spracúvaní obsahu zadania, generovaní odpovedí a výstupov, pričom častokrát dokážu rýchlo ponúknuť lepšie výsledky než ľudia. **Využitie je skutočne širokospektrálne a môže byť úspešné pri chápaní rizík a dôslednom aplikovaní zásad správneho použitia a kontroly**.

Flexibilita a univerzálnosť generatívnych systémov AI otvára možnosti pre široké spektrum ich využitia, napr.:

---

ktoré sú zachytené v neurónových váhach (nie v explicitných gramatických pravidlách). **Emergentným dôsledkom sú tzv. sémantické reprezentácie**, ktoré nie sú formálnou sémantickou analýzou tradičnej lingvistiky, ale naučením sa významov a vzťahov medzi slovami z kontextu. **Veľké jazykové modely, resp. genAI vo všeobecnosti, nevykonávajú syntaktickú a sémantickú analýzu – ony ju „vnímajú“, resp. realizujú prostredníctvom naučených vzorov**. Preto „porozumenie“ jazyka nie je výsledkom formálneho spracovania gramatiky a významu, ale je ovocím pravdepodobnostného učenia na veľkom množstve trénovacích dát.

186 *What is GPT (generative pretrained transformer)?* [on-line]. [cit. 20. septembra 2024]. Dostupné na internete: <https://www.ibm.com/think/topics/gpt>

187 *What is GPT (generative pretrained transformer)?* [on-line]. [cit. 20. septembra 2024]. Dostupné na internete: <https://www.ibm.com/think/topics/gpt>

- chatboty a hlasový asistenti s dotrénovaním pre špecifické oblasti využitia (napr. bankový sektor a finančné poradenstvo, zdravotníctvo, komunikácia špecifických informácií z oblasti činnosti danej firmy či odvetvia priemyslu a pod.);
- vytváranie obsahu a generovanie nielen textu, ale i multimediálneho obsahu;
- jazykové preklady, titulkovanie a tlmočenie;
- sumarizácia a konverzia obsahu;
- analýza a pokročilé spracovanie údajov;
- vytváranie zdrojových kódov a programovanie;
- využitie schopností tzv. „rozumného uvažovania“ sofistikovaných modelov genAI pri riešení zložitých úloh a zadaní;
- nasadenie v oblasti zdravotnej starostlivosti, súdnictva, služieb štátu,....
- samostatné vykonávanie komplexných úloh tzv. agentmi genAI;
- integrácia s inými technológiami a systémami AI (napr. v robotike) pre dosahovanie nových schopností a možností daných technológií, atď.

V zásade možno povedať, že ide o emergentnú a disruptívnu technológiu, ktorá má potenciál dotvárať znalostnú a informačnú spoločnosť a meniť obchodné modely, žurnalistiku, vzdelávanie,[188] sociálne siete a obsah internetu, trh a náplň práce vo viacerých odvetviach. Možno tak hovoriť o potenciálnom dopade na takmer každú oblasť modernej spoločnosti.

Generatívne systémy však so sebou prinášajú aj veľa potencionálnych i reálnych problémov, napr. halucinovanie (vymýšľanie si odpovedí, ktoré systém predkladá ako relevantné a správne), predsudky a neobjektívne výstupy (riziko poloprávd a nesprávnych odpovedí), nejasný spôsob narábania s údajmi (dôsledkom môže byť únik dôverných dát, prehrešky voči ochrane osobných údajov i problémy s autorskými právami), problémy s nahradením niektorých pracovných pozícií, dôsledky pre spoločnosť a demokraciu (propaganda a manipulácia), zneužitie ako nástroj kybernetickej kriminality, riziko digitálnej demencie, erózie identity, deepfake, algoritmického modelovania histórie a pod...[189]

188 Niektoré aspekty dopadu genAI na vzdelávanie sú uvedené v prílohách.

189 Vybraným rizikám generatívnych systémov AI sa budeme venovať v kapitole 2.6.

1.8.1. Ako pracujú generatívne systémy AI?[190]

Veľmi zjednodušene povedané – na implementáciu veľkého jazykového modelu potrebujeme len dva súbory:

- súbor parametrov modelu. Napríklad pre známy opensource model Llama-2-70b[191], ktorý má 70 miliárd parametrov, ide o rozsiahly súbor veľkosti 140GB.[192]
- samotný počítačový program. Ide o krátky súbor cca. 500 riadkov kódu (napr. v jazyku C), ktorý implementuje architektúru neurónovej siete a využíva súbor parametrov pre beh modelu.

Prevádzka modelu (beh modelu) nie je teda nič zložité, len treba pamätať, že pre zodpovedajúci beh a generovanie textu z veľkého jazykového modelu (LLM) treba adekvátny výkon.[193] Algoritmy pre beh LLM sú známe a dostupné, stačí len dostatočný

---

190 *What is GPT (generative pretrained transformer)?* [on-line]. [cit. 20. septembra 2024].
Dostupné na internete: <https://www.ibm.com/think/topics/gpt>
BERMAN, M. *Large Language Models (LLMs) - Everything You NEED To Know.* [on-line]. [cit. 23. september 2024].
Dostupné na internete: <https://www.youtube.com/watch?v=osKyvYJ3PRM>
KARPATHY, A. *Intro to Large Language Models*. [on-line]. [cit. 13. februára 2024].
Dostupné na internete: <https://youtu.be/zjkBMFhNj_g>
Nie je cieľom tejto publikácie podávať technický popis fungovania genAI [napr. detailné vysvetlenie pojmov a technológií transformers, tokenization, encoders, decoders, (un) embedding, (self/multi-head) attention, multilayer perceptrons,...]. Prípadným záujemcom o hlbší technologický pohľad dávame do pozornosti názorne spracovaný seriál videí od 3blue1brown.
3blue1brown: Neural Networks. [on-line]. [cit. 15. novembra 2024].
Dostupné na internete: <https://youtu.be/aircAruvnKk>

191 Llama 2 je open-source kolekcia predtrénovaných a vyladených generatívnych textových modelov od spoločnosti Meta s počtom parametrov od 7 do 70 miliárd.
Llama-2-70b. [on-line]. [cit. 20. septembra 2024].
Dostupné na internete: <https://huggingface.co/meta-llama/Llama-2-70b>

192 Každý parameter je uložený v dvoch bytoch (float16).

193 Napríklad uvedený model Llama-2-70b by bolo problematické prevádzkovať na výkonnom notebooku (treba mať výkonnejšie výpočtové prostredie), ale už model Llama-2-7b (7 miliárd parametrov) by na tom istom notebooku uspokojivo pracoval.
Vývoj však pokračuje – koncom roka 2024 predstavený model Llama-3.3-70b dosahuje výkon porovnateľný s väčším modelom Llama-3.1-405b a je možné ho vyskúšať i na výkonnejších laptopoch. A začiatkom roka 2025 zverejnený open-source čínsky model DeepSeek-R1 s kvalitou na úrovni

výpočtový výkon.

Ak pominieme tŕnistú cestu k vytvoreniu potrebných algoritmov, **tá skutočná „mágia“ je v spôsobe tvorby a získavania súboru parametrov, t.j. v učiacom sa procese**...

Tvorba parametrov predstavuje technologicky a výpočtovo komplexný a náročný proces. Naviac, tie najlepšie súčasné modely (napr. GPT-4o/.../5, Claude 3.7 Sonnet, Gemini 2.5pro + Deep Research, Perplexity AI + DeepResearch, DeepSeek V3/R1, Grok 4 Heavy, ...) obsahujú nie desiatky miliárd (10^9-10^{10}), ale bilióny (10^{12}-10^{13}) parametrov.[194]

Stupne trénovania a prípravy modelu:

0. stupeň: Zhromažďovanie a príprava veľkého množstva dát (**data pre-processing**). V zásade platí, že čím väčší a kvalitnejší model trénujeme (miliardy až bilióny parametrov), tým väčšie množstvo dát je treba použiť na jeho trénovanie.

Dáta by mali byť dostatočne kvalitné, aby bol model relevantne natrénovaný (menej je to dôležité pri tvorbe base modelu, oveľa viac pri ladení a tvorbe assistant modelu – viď ďalšie stupeň trénovania). Rizikom je osobitne nekompletný, nevyvážený a predsudky obsahujúci súbor dát, čo sa následne premietne i do natrénovaného transformačného mechanizmu.

Zdrojom veľkého množstva dát bývajú webové stránky, knihy, odborné publikácie a časopisy, obsah sociálnych sietí a bezplatne prevádzkovaných služieb (napr. emailové služby, multimediálne komunikačné platformy,...), transkripcie textov videí na YouTube,...

Data pre-processing v súčasnosti tvorí celú oblasť dátovej vedy a zahŕňa čokoľvek od riešenia kvality dát, značkovania, konzistentnosti a čistenia dát, ich vyváženosti a normalizácie, transformácie a redukcie, štandardizácie atribútov a pod. Pri veľkých

špičkového OpenAI o1 je možné bez problémov prevádzkovať na výkonnejšom notebooku či PC.
Autor tejto publikácie tiež bežne prevádzkuje viaceré modely (medzi ktoré patrí Llama 3.2, DeepSeek-R1, Mistral, GPT-Oss 5,...) lokálne na svojom notebooku (s využitím techník ako quantization, mixture-of-experts,...).
Informácie o lokálnej prevádzke systémov genAI sú uvedené v prílohe.

194 Odhaduje sa, že už model GPT-4 (po ktorom OpenAI predstavil aj ďalšie modely až po súčasný GPT-5) obsahoval približne 1,8 bilióna parametrov, čo predstavuje viac ako desaťnásobný nárast oproti modelu GPT-3.5.
Dylan Patel, D., Wong, G. *GPT-4 Architecture, Infrastructure, Training Dataset, Costs, Vision, MoE.* [online]. [cit. 10. júla 2023].
Dostupné na internete: <https://semianalysis.com/2023/07/10/gpt-4-architecture-infrastructure/>

súboroch dát ide o časovo a výpočtovo náročný proces (avšak nie viac, ako následné trénovanie základného modelu;-)

1. stupeň: Trénovanie LLM z pripravených trénovacích dát pre generovanie textov je prvým stupňom. Ide o tzv. **predtrénovanie, resp. vytvorenie základného modelu** (base model, foundation model).

V rámci generatívneho predtrénovania ide o trénovanie genAI na neoznačených údajoch, pričom sa model učí rozpoznávať rôzne údaje a zdokonaľuje svoju schopnosť vytvárať presné predpovede. Na základe tohoto trénovania potom generatívne systémy AI pri generovaní výstupov aplikujú vzory a štruktúru vytvorenú predtrénovaním na zadané vstupy (prompty).[195]

Na prvom stupni sa preferuje kvantita pred kvalitou (milióny/miliardy spracovaných dokumentov). Keďže ide o stupeň, ktoré je veľmi výpočtovo, časovo i finančne nákladný, pri skutočne veľkých modeloch sa realizuje približne raz za niekoľko mesiacov až rok.[196]

---

195 Generatívne predtrénovanie je forma učenia bez učiteľa (unsupervised learning), pri ktorom sa modelu poskytujú neoznačené údaje a model je nútený ich sám pochopiť. Tým, že sa modely strojového učenia naučia zisťovať vzory v neoznačených súboroch údajov, získajú schopnosť vyvodzovať podobné závery, keď sú vystavené novým vstupom, ako je napríklad výzva používateľa v aplikácii ChatGPT.
What is GPT (generative pretrained transformer)? [on-line]. [cit. 20. septembra 2024].
Dostupné na internete: <https://www.ibm.com/think/topics/gpt>

196 Napr. prvý stupeň pre trénovanie Llama-2-70b znamená využitie cca. 10TB textu, spracovanie dát na výpočtovom clusteri cca. 6000 GPU, vytvorenie súboru parametrov modelu trénovaním neurónovej siete v priebehu približne dvanástich dní – to všetko v cene asi dvoch miliónov dolárov. Výsledkom je vytvorený tzv. base/foundation model.
Celkové náklady na model Llama-2-70b (nielen base model) boli cca. 3.9 milióna dolárov. Náklady na súčasné modely sa pohybujú v stovkách miliónov dolárov (napr. modely z roku 2023: GPT-4 78,4 milióna USD, Gemini Ultra 191 miliónov USD).
The 2024 AI Index Report. [on-line]. [cit. 5. marca 2025].
Dostupné na internete: <https://hai.stanford.edu/ai-index/2024-ai-index-report>
V súčasnosti možno povedať, že cena implementácie jednoduchých projektov AI sa pohybuje v rozmedzí 5-100 tisíc USD, komplexnejšie projekty medzi 100-500 tisíc, vysoko komplexné sú v rozmedzí 500 tisíc až 1 milión USD a špičkové modely vysoko presahujú 1 milión USD končiac v stovkách miliónov USD. Hardvér sa na celkových nákladoch podieľa v rozmedzí 47-67%, náklady na zamestnancov výskumu a vývoja tvoria 29-49% a zvyšných 2-6% pripadá na spotrebovanú energiu.
Náklady na trénovanie jazykového modelu AI sa medziročne (2023/2024) zvýšili 2,4-krát, pričom je predpoklad, že vývoj špičkových modelov AI by sa časom mohol vyšplhať na 1 miliardu USD. Inovácie v oblasti vývoja AI by sa tak mohli stať nedostupnými pre väčšinu malých a stredných podnikov.

2. stupeň: Ďalším stupňom je **jemné ladenie (fine tuning), ktorého výsledkom je tzv. assistant model**. Ide o dotrénovanie modelu tak, aby nielen generoval text ako pokračovanie textu z promptu, ale aby aj korektne odpovedal na otázky.[197]

Ide o trénovanie ľuďmi, ktorí označujú, identifikujú, resp. vyberajú správne odpovede a prostredníctvom značenia (labelingu) vylepšujú model.[198]

Na druhom stupni sa preferuje kvalita pred kvantitou (počet spracovaných dokumentov je rádovo v stovkách tisíc, pričom však všetky tieto dokumenty by mali byť vysoko kvalitné konverzačné dokumenty, ktoré podľa zadaných inštrukcií pre značenie vytvárajú ľudia).

Asistent model na 2. stupni potom odpovedá nielen na základe naučených informácií z 1. stupňa, ale naviac pridáva tzv. *alignment* a formátovanie, resp. správny spôsob odpovede na položenú otázku.

Zjednodušene môžeme povedať, že prvý stupeň učenia je o dátach a informáciách, druhý o spôsobe ich využitia. V druhom stupni tiež môžeme model prispôsobovať (customizovať), t.j. pridávať dotrénovanie pre špecifické použitie, napr. pre robotiku, finančné poradenstvo, analytické nástroje a pod.

Druhý stupeň zahŕňa:

- spísanie inštrukcií značenia (labeling instruction), ktoré špecifikujú, ako sa model bude správať;
- zamestnanie ľudských zdrojov (resp. použitie systémov AI, napr. scale.ai), ktoré pracujú na zhromažďovaní stoviek tisíc vysoko kvalitných textov s otázkami a odpoveďami, resp. porovnaniami odpovedí;
- vyladenie base modelu na takto pripravených dátach (ladenie prebieha v jednotkách dní);
- vykonanie mnohých kontrol a vyhodnotení kvality;
- nasadenie do prevádzky;

How Much Does AI Cost: A Financial Analysis for C-Level Executives. [on-line]. [cit. 5. marca 2025]. Dostupné na internete: <https://vodworks.com/blogs/how-much-does-ai-cost/>

197 Base model nie je reálne využiteľný (nenaučil sa odpovedať, len generuje text, takže na zadaný text by len pokračoval generovaním ďalšieho textu).

198 Na 2. stupni sú aplikované predovšetkým algoritmy učenia s učiteľom (supervised learning).

- monitorovanie činnosti modelu v prevádzke, zbieranie chybných odpovedí a zlyhaní, ktoré sa následne opravujú a opakuje sa celý proces druhého stupňa (napr. ak na nejakú otázku dá asistent model nesprávnu odpoveď, vytvorí sa odpoveď nová a správna, ktorá sa následne zahrnie do učenia/ladenia).

Keďže v druhom stupni ide o oveľa lacnejší a rýchlejší proces, môže byť realizovaný raz týždenne alebo priebežne.

3. stupeň: Voliteľný stupeň, v ktorom sa označovanie (labeling) koná nie na základe vytvárania správnych odpovedí, ale na základe porovnávania jednotlivých odpovedí. Využíva sa **reinforcement learning from human feedback** (RLHF, v terminológii OpenAI *early chat*) s využitím porovnávacieho označovania.[199]

V rámci RLHF sa porovnáva kvalita a správnosť (väčšinou i etika) viacerých odpovedí, ktoré na danú otázku generuje 2. stupeň (assistant model). Ide o proces, ktorý je oveľa jednoduchší a rýchlejší, než ďalšie dotrénovanie na vytvorených ideálnych odpovediach.

RLHF prináša oveľa väčšiu presnosť a v kombinácii s prispôsobovaním (customizáciou) sa model adaptuje na reálne oblasti využitia. RLHF a prispôsobovanie modelu je možné vykonávať opakovane a pre rôzne scenáre a vytvárať tak rôzne služby postavené na generatívnych systémoch AI.

Nakoľko RLHF vyžaduje intenzívne, odborné, vyvážené a eticky korektné zapojenie trénovacieho tímu, viacerí tvorcovia veľkých systémov genAI (Anthropic, OpenAI, Google,...) na 3. stupni využívajú i **reinforcement learning from AI feedback** (RLAF).[200]

---

199 3. stupeň využíva implementáciu učenie formou odmeňovania (reinforcement learning).
V oblasti reinforcement learningu pre LLM/VLM prebieha veľmi intenzívny vývoj a okrem RHLF prichádzajú na scénu napr. assistance game, MCTS modely pre implementáciu uvažovania (simuláciu myslenia typu System2 – viď ďalšia kapitola) a pod.
MURPHY, P. K. *Reinforcement Learning: An Overview*. [on-line]. [cit. 9. decembra 2024]. Dostupné na internete: <https://arxiv.org/abs/2412.05265>

200 RLAF (pôvodne vyvinutý firmou Antropic, ktorá stojí za genAI Claude) nahrádza ľudských anotátorov systémom AI, ktorý poskytuje signály odmeny, resp. preferenčné hodnotenia. **V 3. stupni sa pri RLAF začleňuje spätná väzba od iného AI modelu, nie od ľudí**. Ľudský faktor sa podieľa na vytvorení tzv. constitution (v terminológii RLAF), t.j. základných pravidiel, podľa ktorých anotujúci model AI robí svoje rozhodnutia a dolaďuje trénovaný veľký model genAI.
Výhodou je masívne škálovanie, zníženie nákladov, konzistentnosť činnosti, implementácia slučiek seba-zdokonaľovania a pod.
K rizikám patrí možnosť zaujatosti, kompromitácia (hackovanie) odmien, nadmerná závislosť

Technológie tvoriace generatívne systémy AI:[201]

- **tokenization** – ide o techniku delenia dlhého textu na tokeny, ktoré sú prevažne tri až štyri znaky dlhé. Tokeny sú tvorené časťami slov, reťazcami znakov alebo kombináciami krátkych slov, znakov a interpunkčných znamienok, ktoré sa využívajú na rozklad textu. Tokenizácia je prvým krokom pri trénovaní, pričom sa tento proces môže líšiť model od modelu.
- **embedding** – vkladanie tokenov do vektorov na prevedenie tokenov do hromady ich numerických reprezentácií, pričom každý token korešponduje s konkrétnou pozíciou v databáze vektorov. Pre počítač to predstavuje nielen význam daného tokenu, ale aj jeho vzťah k ostatným tokenom. V reále môže ísť o mnohorozmernú databázu vektorov. V rámci embeddingu sa realizuje proces mapovania tokenov na virtuálny trojrozmerný vektorový priestor, pričom sa predpokladá, že tokeny zakódované blízko seba v trojrozmernom priestore sú si významovo podobnejšie. Moduly realizujúce túto matematickú vektorizáciu vstupnej sekvencie sa nazývajú enkodery (**encoders**).[202]

---

od procesov a vlastností použitého systému AI (zmierňuje sa to ľudskými kontrolami a bezpečnostnými obmedzeniami).

O'CONNOR, R. *How Reinforcement Learning from AI Feedback works*. [on-line]. [cit. 29. augusta 2025]. Dostupné na internete: <https://www.assemblyai.com/blog/how-reinforcement-learning-from-ai-feedback-works>

201 Všeobecným základom pre genAI sú **tri základné myšlienky, ktoré znamenali veľký posun v oblasti vývoja súčasných systémov umelej inteligencie**:

– využitie autoregresívnych modelov trénovaných na textoch;

– realizácia skutočne veľkých neurónových sietí s dostatočným množstvom parametrov;

– trénovanie na obrovských súboroch dát.

SUTSKEVER, I., VINYALS, O., LE, Q.V. *Sequence to Sequence Learning with Neural Networks*. [on-line]. ArXiv, abs/1409.3215, 2014. [cit. 13. decembra 2024].

Dostupné na internete:

<https://www.semanticscholar.org/reader/cea967b59209c6be22829699f05b8b1ac4dc092d>

202 Enkodery v sieti transformerov priraďujú každému embeddingu váhu, ktorá určuje jeho relatívnu dôležitosť. Medzitým pozičné enkodery zachytávajú sémantiku, čo umožňuje modelom genAI rozlišovať medzi zoskupeniami rovnakých slov, ale v rôznom poradí - napríklad „Vajíčko prišlo pred sliepkou“ v porovnaní so „Sliepka prišla pred vajíčkom“.

What is GPT (generative pretrained transformer)? [on-line]. [cit. 20. septembra 2024].

Dostupné na internete: <https://www.ibm.com/think/topics/gpt>

- **transformers** – z týchto vektorov je možné vytvoriť maticové reprezentácie. Prostredníctvom algoritmu nazývaného **self-attention**, resp. **multi-head attention** sa z čísel vyextrahujú niektoré informácie a umiestnia sa do matice. Výstupom tohto algoritmu je súbor čísel v matici – matematická reprezentácia, ktorá modelu hovorí, ako veľmi sa slová a ich poradie podieľajú na vete ako celku.[203]

  Pri generovaní vstupnú maticu transformujeme na výstupnú maticu, ktorú potom prevedieme do prirodzeného jazyka pomocou dekodérov (**decoders**).[204] A slovo je konečným výstupom celého tohto procesu (samozrejme v prípade multimodálnych genAI je výstup nielen textový).

  Túto transformáciu vykonáva algoritmus, ktorý bol spolu s parametrami vytvorený počas procesu trénovania. Schopnosť modelu vykonať túto transformáciu je teda založená na všetkých jeho znalostiach, ktoré boli získané z trénovacích dát. Model sa na základe váh určených počas trénovania naučil, ktoré sekvencie slov idú spolu a aké ďalšie slová s najväčšou pravdepodobnosťou nasledujú.[205]

Modely založené na transformeroch jazyk nespracúvajú a nerozumejú mu rovnakým spôsobom ako ľudia. Narábajú s tokenmi a tým, že vyhodnocujú všetky tokeny naraz, vynikajú v určovaní závislostí s veľkým dosahom, t.j. v určovaní vzťahov medzi vzdialenými tokenmi. Generatívne systémy AI sa pri kontextovom spracovaní vstupov spoliehajú na svoje chápanie závislostí s veľkým dosahom (v rámci kontextového okna).

---

203 Transformátory používajú mechanizmus na pochopenie kontextu slov vo vnútri vety. Tento mechanizmus zahŕňa výpočty s bodovým súčinom, čo je v podstate číslo vyjadrujúce, ako veľmi slovo prispelo k vete. Mechanizmus zistí rozdiel medzi bodovými súčinmi slov a priradí im zodpovedajúco veľké hodnoty pozornosti. A ak má slovo vyššiu pozornosť (attention), bude sa naň prihliadať viac. Bodový súčin alebo skalárny súčin dvoch vektorov je spôsob násobenia dvoch vektorov. Z geometrického uhla pohľadu je bodový súčin súčinom dĺžky vektorov a kosínusu uhla medzi nimi.

204 Dekodéry predpovedajú štatisticky najpravdepodobnejšiu odpoveď na vektorizáciu vstupnej sekvencie pripravenú enkodermi. Mechanizmy self-attention a multi-head self-attention umožňujú dekodérom identifikovať najdôležitejšie časti vstupnej sekvencie, zatiaľ čo pokročilé algoritmy určujú výstup, ktorý je s najväčšou pravdepodobnosťou správny.
What is GPT (generative pretrained transformer)? [on-line]. [cit. 20. septembra 2024]. Dostupné na internete: <https://www.ibm.com/think/topics/gpt>

205 Transformery sú teda typom neurónovej siete špecializovanej na spracovanie prirodzeného jazyka identifikujúc zámer a význam v textovom vstupe. Dokážu dynamicky spracovávať vstupy a zamerať sa na najdôležitejšie slová bez ohľadu na to, kde sa vo vete nachádzajú.

Natrénované modely rozlišujú syntaktické vzory a vzťahy, ktoré sú zachytené v neurónových váhach (nie v explicitných gramatických pravidlách). Emergentným dôsledkom sú tzv. sémantické reprezentácie, ktoré nie sú formálnou sémantickou analýzou tradičnej lingvistiky, ale naučením sa významov a vzťahov medzi slovami z kontextu. Veľké jazykové modely, resp. systémy genAI vo všeobecnosti, nevykonávajú syntaktickú a sémantickú analýzu – ony ju „vnímajú“, resp. realizujú prostredníctvom naučených vzorov. Preto **„porozumenie“ jazyka nie je výsledkom formálneho spracovania gramatiky a významu, ale je ovocím pravdepodobnostného učenia na veľkom množstve trénovacích dát**.[206]

1.8.2. Vylepšovanie generatívnych systémov a budúci vývoj

Oblasť generatívnych systémov AI v súčasnosti zažíva veľký progres, ktorého obsahom je neustále zlepšovanie výkonu a schopností genAI modelov: zväčšovanie kontextového okna, lepšie chápanie kontextu, zlepšovanie presnosti a relevantnosti výsledkov, generovanie presnejších a konzistentnejších výstupov, spoľahlivejšie odpovede v bežných konverzáciách, realizácia schopností logickej dedukcie a komplexnej syntézy informácií, reálne uplatnenie v pracovných úlohách (napr. kódovanie, analýza dát, príprava zmlúv, výskum,...), sumarizácia zdrojov a tvorba rešerší, riešenie komplexných úloh, spresnenie numerických výpočtov, jazyková koherencia pri generovaných dlhších výstupoch, atď.

Osobitnou a dôležitou oblasťou vývoja je odolnosť (spoľahlivosť a bezpečnosť), hodnotové nastavenie a etika, nakoľko veľkí hráči na poli vývoja špičkových systémov AI čelia zo strany viacerých odborníkov kritike, že vývoj ide prirýchlo a bez dostatočnej verejnej kontroly. **Problematika tzv. „alignmentu“, t.j. zosúladenie správania systémov AI s ľudskými hodnotami a cieľmi, patrí podľa nášho názoru k podstatným aspektom vývoja a pokroku v oblasti AI**.[207]

Keďže vývoj v oblasti generatívnych systémov AI v súčasnosti zažíva neskutočný

206 **Generatívne modely AI postavené na transformeroch teda spracúvajú vstupné údaje rozdelené na tokeny a vektorizované technológiou embeddingu v rámci enkodera, pričom na stanovenie závislostí a vzťahov používajú mechanizmy pozornosti (self-attention, multi-head self-attention). Takto vytvorená vstupná matica sa následne transformuje na výstupnú maticu a jej prevedením na text (resp. žiadanú modalitu) sa generuje výstup prostredníctvom dekodera. Spôsob transformácie je podmienený natrénovanými parametrami systému**.

207 Ide o problematiku tzv. dôveryhodných systémov AI, t.j. na dobro človeka orientovaných systémov AI, ktorej sa venujeme v 4. kapitole.

rozmach, asi nie je možné podchytiť všetky aspekty technologického rozvoja a neustály prúd vedeckých prác i praktických aplikácií v oblasti primárneho a aplikovaného výskumu i aktuálnych trendov. Uvádzame aspoň niekoľko základných poznatkov, trendov a perspektívnych smerov vývoja.

Možnosti škálovania generatívnych systémov AI[208]

V súčasnosti je výkon LLMs jasne predikovateľná funkcia dvoch parametrov:[209]

- N, počet parametrov v neurónovej sieti.
- D, množstvo textu, resp. vstupných dát, na ktorých trénujeme.

Obrazne povedané, „zadarmo“ môžeme získať „viac inteligencie“ len škálovaním generatívneho systému bez potreby zlepšovania algoritmov. Týmto spôsobom sa môže realizovať nielen zvyšovanie presnosti predikcie ďalšieho slova, ale aj zvyšovanie všeobecných schopností naprieč všetkými oblasťami využitia.

Ideálnym spôsobom zvyšovania výkonu a zlepšovania kvality systémov genAI by malo byť vylepšovanie modelov a trénovacích datasetov i hľadanie nových algoritmov, čo určite nie je triviálna záležitosť.[210]

Okrem vylepšovania a hľadania nových algoritmov (čo je určite správne) je teda možné vylepšovať generatívne systémy AI aj zvyšovaním výkonu existujúcich systémov prostredníctvom zvyšovania počtu parametrov neurónovej siete a množstva vstupných

---

208 KARPATHY, A. *Intro to Large Language Models*. [on-line]. [cit. 13. februára 2024]. Dostupné na internete: <https://youtu.be/zjkBMFhNj_g>

209 Niekedy sa uvádza i veľkosť kontextového okna, prípadne ďalšie technologické parametre, tie však podľa nášho názoru patria skôr k sekundárnym aspektom závislým či už od uvádzaných základných parametrov alebo od ďalšieho narábania s prvotnými base a assistent modelmi. Poniektorým z týchto aspektov sa budeme venovať v ďalšom texte.

210 Že je to možné, dosvedčuje aj zverejnenie čínskeho modelu DeepSeek-V3 a predovšetkým DeepSeek-R1 počiatkom roku 2025. Ide o modely, ktoré sú výkonovo na úrovni špičkových modelov OpenAI o1, Gemini 2.0 a Claude 3.5, pričom technologické nároky na trénovanie i prevádzku sú rádovo nižšie ako u uvedených modelov. Keďže v prípade DeepSeek-R1 ide o open-source model, má zverejnený aj zdrojový kód, parametre a rozsiahlu dokumentáciu. **Javí sa, že v prípade V1/R1 ide o seriózne vylepšenia, ktoré umožnili týmto modelom vysoký výkon a výborné výsledky v testoch na úrovni tých najlepších modelov**.
DeepSeek-R1 Release. [on-line]. [cit. 25. januára 2025]. Dostupné na internete: <https://api-docs.deepseek.com/news/news250120>

trénovacích dát (vo všeobecnosti kvalita a množstvo trénovacích dát sú nutnou podmienkou akéhokoľvek špičkového generatívneho systému súčasnosti).

Používanie externých nástrojov jazykovými modelmi

Existujú mnohé vstupné zadania (prompty), na ktoré generatívne systémy nevedia uspokojivo odpovedať.[211] Jedným z riešení hľadania správnych odpovedí je i prepájanie systémov genAI s externými nástrojmi.[212]

Napríklad v prípade vstupného zadania (promptu), u ktorého systém genAI zistí, že ho nevie zodpovedať, použije externé nástroje na vykonanie úlohy:

- využitie vyhľadávača/jeho databázy na získanie potrebných podkladov. LLM potom na základe získaných podkladov vo forme vyhľadaného textu generuje odpoveď.
- na vykonanie výpočtového zadania (na základe špeciálnych slov v prompte, resp. prvotného spracovania zadania) použije kalkulačku alebo matematické softvérové knižnice.
- na vizualizáciu výsledkov podľa zadania promptu (napr. vykreslenie grafu podľa konkrétnych špecifikácií) využije vhodné softvérové knižnice (napr. matlab lib, gnuplot,...).
- z takto vytvorenej bázy dát odpovedí je možné vytvoriť zadanie pre iný model, napr. obrazový, video a pod...

Pre efektívne využívanie nástrojov sa vyžadujú určité analytické schopnosti a pokročilá úroveň daného generatívneho systému AI. **Využívanie externých nástrojov v rámci existujúcej výpočtovej infraštruktúry patrí k podstatným aspektom vzrastajúcich schopností systémov genAI.** Využívanie nástrojov sa v hojnej miere využíva i v rámci ďalších technológií, ktoré popisujeme v tejto kapitole (napr. komplexné rozhodovanie, RAG, genAI agenti, genAI ako operačný systém a pod.).

---

211 Klasické LLM sú napríklad pomerne slabé v matematike, logike a ďalších deterministických záležitostiach.

212 Na bezproblémovú integráciu medzi aplikáciami LLM a externými zdrojmi údajov sa využíva tzv. **Model Context Protocol**. Ide o otvorený štandard, resp. protokol vyvinutý spoločnosťou Anthropix, ktorý bol navrhnutý tak, aby poskytoval univerzálny spôsob prepojenia externých nástrojov a zdrojov údajov so systémami AI.
Model Context Protocol. [on-line]. [cit. 22. marca 2025].
Dostupné na internete: <https://github.com/modelcontextprotocol>

V rámci používania externých nástrojov pre riešenie sofistikovaných zadaní systémy genAI využívajú zapojenie rôznych nástrojov, čo sa veľmi podobá aj ľudskému postupu riešenia problémov. Ide však stále len o simuláciu ľudského prístupu v rámci systému ANI, nakoľko toto využitie je algoritmizované a natrénované.

Určitou modifikáciou tohto prístupu je i pomerne nová technológia, ktorá umožňuje akési spojenie viacerých modelov. Jednotlivé modely sú navrhnuté a vytrénované tak, aby boli odborníkmi v určitých oblastiach, a následne, keď príde skutočná výzva, systém vyberie, ktorý z týchto expertných modelov použije.

Pokrok v rámci multimodálnych modelov LMM/VLM

V rámci multimodálnych generatívnych systémov AI môžu byť rôzne modality vstupu (obraz, video, audio,...) nielen generované, ale aj analyzované a ďalej spracúvané, resp. kombinované. Môže ísť o prijímanie textových a audio vstupov, obrázkov, videa a všetkých možných vstupných zdrojov a **ich jednotné spracovanie a výstup**. Napríklad:

- na základe obrazového vstupu, ktorý sa zanalyzuje, systém vytvára odpoveď a ak treba, znovu ju i graficky vyjadrí a vygeneruje obrázok.
- celá komunikácia s modelom je hovoreným slovom, t.j. ako prompt poviem otázku a ako odpoveď sa generuje hovorené slovo.[213]

Súčasťou pokroku vo vývoji multimodálnych modelov je rozvoj ich schopností v oblasti spracovania prirodzeného jazyka: kontextovo orientovaný prepis a spracovanie reči v rôznych jazykoch, generovanie syntetických hlasových výstupov takmer nerozoznateľných od hlasu človeka, pokročilé hlasové schopnosti (prirodzená konverzácia s reálnou interpunkciou, akcentom, prízvukom...), výborná replikácia a napodobňovanie

213 Túto funkcionalitu ponúka napr. GPT-4o v rámci tzv. rozšíreného hlasového režimu (Advance Voice Mode). Snahou OpenAI je čo najviac priblížiť hlasové konverzácie s GPT-4o rozhovorom s ľudskou bytosťou: chatbot ponúka vylepšené reakcie (napr. reakcia počas odpovedania na prerušenie používateľom, ktorý chce doplniť doplňujúcu otázku), dokáže používať určité emócie (má vedieť reagovať na humor), má funkciu pamäte a mal by vedieť využívať aj neverbálne indície (napr. rýchlosť reči) a pod.
Hello GPT-4o. [on-line]. [cit. 19. augusta 2024].
Dostupné na internete: <https://openai.com/index/hello-gpt-4o/>
Introducing GPT-4o, our new model which can reason across text, audio, and video in real time. [on-line]. [cit. 20. augusta 2024].
Dostupné na internete: <https://twitter.com/gdb/status/1790071008499544518>

ľudského hlasu z niekoľkominútových vzoriek a pod.

Najnovšie systémy genAI dokážu priamo spracovávať rôzne modality vstupu bez potreby ich konverzie (príklad potreby konverzie: vstupný obrázok sa prevedie na popisný text, na základe ktorého sa generuje výstupný text, ktorý sa následne znovu prevedie na obrázok alebo audio výstup).[214]

Na prelome rokov 2024/2025 bolo zverejnených viacero multimodálnych modelov, ktoré dosahujú vynikajúce vlastnosti nielen pri generovaní multimediálnych výstupov, ale predovšetkým v schopnosti integrovať rôzne modality vstupu, v pokročilom spracovaní a prepojení s veľkými jazykovými modelmi i v generovaní praktických multimediálnych výstupov. Veľký pokrok nastal aj v schopnosti niektorých z nich interaktívne komunikovať s používateľom na základe multimodálnych vstupov, napr. mobilom snímanej scény a k tomu hovoreného slova používateľa mobilu. Medzi tieto modely patrí napr. Google Astra, Grok 2 + Aurora, OpenAI Sora, Copilot Vision, Gemini 2.0 + Mediapipe, atď. (a samozrejme aj ich novšie verzie:-)

Komplexné rozhodovanie generatívnych systémov[215]

Daniel Kahneman vo svojej knihe Thinking, Fast and Slow[216] porovnáva dva spôsoby myslenia – System1 a System2:

- System1: podvedomé, automatické, rýchle rozmýšľanie, mnohokrát emočné a bez námahy (napr. 2+2).
- System2: vedomé, racionálne, pomalšie, komplexné rozhodovanie, viac logické, s vynaloženým úsilím (napr. 17*24).

Napríklad v šachu by System1 generoval návrhy na ďalší ťah ako pri tzv. rýchlom šachu (okamžité ťahy). System2 by však sledoval strom možných ťahov, rozmýšľal a zvažoval množstvo možností a potencionálnych riešení.

Doterajšie generatívne systémy AI pracovali ako System1 – ide o reťaz jednoduchých krokov, ktoré však nevedia dať správne odpovede na zložitejšie

214 GPT-4o je prvý tzv. **omni-modálny systém**, ktorý umožňuje natívne spracovávať textové, obrazové, video a audio vstupy, pričom dokáže generovať textové, obrazové a audio výstupy.

215 KARPATHY, A. *Intro to Large Language Models*. [on-line]. [cit. 13. februára 2024]. Dostupné na internete: <https://youtu.be/zjkBMFhNj_g>

216 KAHNEMAN, D. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2011. ISBN 978-0374275631.

otázky.

Preto v súčasnosti prichádzajú riešenia, ktoré umožňujú systémom genAI simulovať myslenie System2. Snahou je na úkor rýchlosti smerovať k presnosti, t.j. od priamočiareho generovania odpovede prejsť k hľadaniu optimálnej odpovede premýšľaním o problémoch krok za krokom a výberom v strome možností (ToT – tree of thoughts, CoT – chain of thoughts).

Ak by sme uvažovali o závislosti presnosti generovaných odpovedí od času, ktorý má systém na zodpovedanie promptu, pri System1 by presnosť bola nezávisle od času stále rovnaká, no pri System2 by presnosť s časom rástla (funkcionál by samozrejme mohol vyzerať rôzne, podľa zvoleného riešenia, no časom by konvergoval k maximálnej možnej presnosti v rámci daného modelu a algoritmov).[217]

Simuláciu myslenia System2 do určitej miery implementujú najnovšie verzie niektorých generatívnych systémov AI, pričom vývoj rýchlo napreduje. Napríklad v rámci modelu o1 od spoločnosti OpenAI, ktorý sa snaží komplexné uvažovanie analogické System2 implementovať, je vidieť evidentný posun oproti len o niekoľko mesiacov staršiemu modelu GPT-4o – model o1 ponúka vyššiu presnosť logického uvažovania, lepšie porozumenie dlhším textovým kontextom, zníženú mieru nepresností a mal by dosahovať lepšie výsledky v náročnejších scenároch, ako je tvorba detailných rešerší, spracovanie odborných dokumentov či analýza zložitých otázok.[218]

217 Presnosť s časom pri System2 rastie v ideálnom prípade, resp. pri správnej implementácii tohto spôsobu „myslenia“. Z reálneho nasadenia súčasných uvažujúcich genAI poznáme prípady, keď hľadanie optimálnej odpovede vedie k zavrhnutiu správneho riešenia a postupnej konvergencii k riešeniu horšiemu.
DZIRI, N. *We noticed that many failures stem not from lack of knowledge but from overthinking*. [on-line]. [cit. 25. júna 2025].
Dostupné na internete: <https://x.com/nouhadziri/status/1937567619051049446>
SUN, Y. HU, S., ZHOU, G. et al. *OMEGA: Can LLMs Reason Outside the Box in Math? Evaluating Exploratory, Compositional, and Transformative Generalization*. [on-line]. [cit. 25. júna 2025].
Dostupné na internete: <https://doi.org/10.48550/arXiv.2506.18880>

218 *Hello GPT-4o*. [on-line]. [cit. 19. augusta 2024].
Dostupné na internete: <https://openai.com/index/hello-gpt-4o/>
Learning to Reason with LLMs. [on-line]. [cit. 20. novembra 2024].
Dostupné na internete: <https://openai.com/index/learning-to-reason-with-llms/>
Produkčné verzie modelov o1 a o1 Pro od OpenAI (december 2024) dosahujú skutočne pozoruhodné výsledky. Napríklad veľmi presné vykonávanie inštrukcií z promptov, veľké zlepšenie v generovaní

Nástupcom modelu o1 je model o3-mini, ktorý OpenAI oznámila koncom roku 2024. S týmto modelom a jeho analógiami v podobe najnovších verzií Claude 3.7 Sonnet, Gemini 2/2 pro + AI co-scientist, Perplexity AI + DeepResearch, DeepSeek R1 a viacerými ďalšími, ktoré boli predstavené v priebehu roku 2025 a ktoré implementujú tzv. rozumné uvažovanie (common-sense reasoning), prichádza tzv. éra uvažovania (**reasoning era**), znamenajúca snahu o implementáciu myslenia System2 a realizáciu systémov BAGI (Below-human AGI), ktoré sú viacerými odborníkmi uvádzané ako technológie na pomedzí úzkych a slabých systémov (ANI) a všeobecných a silných systémov umelej inteligencie (AGI).[219] [220]

Modely, ktoré implementujú tzv. rozumné uvažovanie, dosahujú veľmi dobré výsledky v chápaní kontextu a tvorivom riešení problémov. Pred poskytnutím odpovede venujú viac času podrobnej analýze problémov, plánovaniu riešení a zvažovaniu viacerých perspektív.

Ak by sme nadviazali na debatu z kapitoly 1.7. ohľadom hľadania nových riešení smerujúcich k všeobecnej umelej inteligencii (AGI) kombináciou technológií symbolickej a subsymbolickej AI, **implementácia System2 by mohla byť súčasťou prekročenia subsymbolického prístupu rýchleho vnímania (fast perception)**. Otázku, či skutočne pôjde aj o implementáciu a integráciu symbolických postupov (napr. logiky prvého rádu), musia zodpovedať vývojové tímy tých najlepších systémov genAI. Ak by k tomu skutočne prišlo, možno by to malo potenciál vykonať ďalší krok na ceste k skutočnej umelej inteligencii.

Treba však zdôrazniť, že **určite nezdieľame terajšie nadšenie, ktoré považuje súčasné uvažujúce modely za začiatok éry AGI**, t.j. skutočnej silnej a všeobecnej umelej inteligencie (nezávisle od úrovne, t.j. BAGI, HLAGI,...).

---

počítačového kódu, chápanie obsahu výkresov, schopnosť riešiť neľahké zadania, vylepšená schopnosť tvorby grafických výstupov, ale aj využitie vlastného modelu na ochranu pred zneužitím a ohrozením, no i snahu o zvrátenie dohľadu a klamanie používateľa.

WILLISON, S. *Here's the spiciest detail from the new o1 system card*. [on-line]. [cit. 10. decembra 2024]. Dostupné na internete: <https://x.com/simonw/status/1864737207111815177>

219 *Early access for safety testing*. [on-line]. [cit. 22. decembra 2024]. Dostupné na internete: <https://openai.com/index/early-access-for-safety-testing/>

220 GABE, M. *Four Phases of AGI*. [on-line]. [cit. 22. decembra 2024]. Dostupné na internete: <https://www.lesswrong.com/posts/qeJomTN2yp5tQG4rL/four-phases-of-agi>

I napriek neuveriteľným možnostiam, ktoré uvažujúce genAI systémy ponúkajú, stále vykazujú veľké problémy v schopnosti skutočne riešiť celú škálu úloh (z ktorých niektoré nie je problém zvládnuť inými algoritmami AI). Majú veľké problémy s relevantnosťou a správnosťou ponúkaných riešení, vykazujú dramatický pokles výkonu v závislosti na zložitosti riešených úloh, nedokážu spoľahlivo vykonávať algoritmy (napr. nemajú internalizované základné algoritmy, ale všetko riešia na základe vzorov), v rámci CoT (chain of thoughts, System2) sú náchylné na tzv. overthinking, čiže ďalším premýšľaním o zadanom probléme zavrhnú správnu odpoveď z rannej fázy CoT a v rámci začarovaného kruhu samokorekcií správne riešenie opustia, s narastajúcou zložitosťou problémov uvažujúce modely môžu premýšľať menej a horšie, nie viac a lepšie, atď.[221]

Niektoré problémy súčasných modelov rozumného uvažovania nevyplývajú z ich nedostatočnej kvality, ale z nemožnosti kontrolovať svoju činnosť po jednotlivých krokoch, resp. uprostred uvažovania. Modely jednoducho nemajú implementovanú vnútornú kontrolu a kritiku, nemajú implementovaný spôsob medzistupňového odmeňovania, napr. formou uvažovania o jednotlivých medzistupňoch riešenia, ktoré vykonáva tzv. reasoning model. **V zásade učíme systémy genAI na základe vzorov riešiť problémy bez toho, že by sme implementovali ich „vnútorný monológ“, meta-uvažovanie, ktoré kriticky vyhodnocuje každý krok riešenia zadaného problému**.[222]

I v optike dosiahnutia méty AGI je chýbajúca metakognícia obrovským problémom súčasných uvažujúcich genAI: veľké uvažujúce modely (LRM) síce vykazujú sofistikované porovnávanie vzorov a dokážu generovať prepracované stopy uvažovania, ale v podstate im chýba metakognitívne vedomie vyššieho rádu, ktoré by im umožňovalo monitorovať, kalibrovať a prispôsobovať svoje vlastné procesy uvažovania. Takže LRMs si svoje procesy neuvedomujú a nevedia ich ani primeranie (t.j. vedome, alebo minimálne správne a skutočne inteligentne) prispôsobovať.[223]

221 SUN, Y. HU, S., ZHOU, G. et al. *OMEGA: Can LLMs Reason Outside the Box in Math? Evaluating Exploratory, Compositional, and Transformative Generalization*.
PEREZ, C.E. *Looking at the "Illusion of Thinking"*. [on-line]. [cit. 25. júna 2025].
Dostupné na internete: <https://x.com/IntuitMachine/status/1931644650894340098>

222 XIONG, W., ZHAO, W., YUAN, W. et al. *StepWiser: Stepwise Generative Judges for Wiser Reasoning*. [on-line]. [cit. 24. septembra 2025].
Dostupné na internete: <https://arxiv.org/pdf/2508.19229>

223 Pohľad na „ilúziu myslenia“ cez optiku metakognitívneho rámca QPT odhaľuje **hlbokú slabinu epistemického sebauvedomenia veľkých modelov uvažovania (LRM - large reasoning model) –**

Vzhľadom na uvedené skutočnosti môžeme súhlasiť s tvrdením, že súčasné veľké uvažujúce modely genAI nás ku skutočnej umelej inteligencii (AGI) už z princípu neprivedú.[224] Ich schopnosti – rozpoznávanie vzorov s ich aplikáciou na nové vstupy a architektúra transformerov, ktorá umožňuje modelu paralelne spracovať všetky časti vstupnej sekvencie a generovať výstupy – znamenajú nepochybne veľký pokrok vo vývoji systémov AI a určite prispievajú i na ceste k AGI, avšak stále neprekračujú koncepčné prelomy, ktoré treba zdolať na ceste ku skutočnej umelej inteligencii.[225]

Hybridné modely uvažovania (hybrid reasoning models)

24. februára 2025 bol predstavený model Claude 3.7 Sonnet od spoločnosti Anthropic, ktorý bol prezentovaný ako „prvý hybridný model uvažovania na trhu“.[226]

Spoločnosť Anthropic pojem „hybridný“ chápe ako schopnosť modelu vytvárať buď rýchle odpovede alebo generovať odpovede krok po kroku na základe „rozšíreného“ uvažovania, vďaka čomu dokáže riešiť i zložité problémy.

Na rozdiel od iných riešení, ktoré implementujú využitie viacerých modelov – jeden pre rýchle odpovede, iný pre zapojenie postupného uvažovania, v prípade Claude 3.7 Sonnet ide o jeden model, ktorý je v štandardnom režime vylepšenou verziou pokročilého systému Claude 3.5 Sonnet, avšak v režime rozšíreného myslenia venuje Claude 3.7 Sonnet viac času podrobnej analýze problémov, plánovaniu riešení a zvažovaniu viacerých perspektív pred poskytnutím odpovede.[227]

---

ich schopnosť vedieť, čo nevedia, a podľa toho kalibrovať svoje kognitívne úsilie.
PEREZ, C.E. *Looking at the "Illusion of Thinking"*.
SHOJAEE, P. MIRZADEH, I., ALIZADEH, K. et al. *The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity*. [on-line]. Apple, 2025. [cit. 25. júna 2025].
Dostupné na internete: <https://machinelearning.apple.com/research/illusion-of-thinking>

224 MARCUS, G. *No way are “reasoning” models like o3 going to get us to AGI*. [on-line]. [cit. 25. júna 2025].
Dostupné na internete: <https://x.com/garymarcus/status/1937612648750711280>

225 Koncepčné prelomy na ceste k AGI diskutujeme v kapitole 5.3.1.

226 *Claude 3.7 Sonnet and Claude Code*. [on-line]. [cit. 7. marca 2025].
Dostupné na internete: <https://www.anthropic.com/news/claude-3-7-sonnet>

227 *Anthropic’s Claude 3.7 Sonnet hybrid reasoning model is now available in Amazon Bedrock*. [on-line]. [cit. 7. marca 2025].
Dostupné na internete: <https://aws.amazon.com/blogs/aws/anthropics-claude-3-7-sonnet-the-first-

V kontexte tejto kapitoly a aktuálneho rozvoja genAI však **myšlienka „hybridného“ systému môže byť rozvíjaná oveľa viac**. Môžeme uvažovať nielen o spojení klasického a uvažujúceho systému AI, ale aj o kombinácii viacerých typov pokročilého uvažovania (nielen klasické delenie na uvažovanie deduktívne, induktívne, abduktívne, atď., ale i viaceré technologické postupy genAI),[228] už spomínanú kombináciu subsymbolického a symbolického prístupu (napr. kombinácia metód hlbokého učenia s logikou prvého rádu) alebo hybridné riešenie problému s rozlišovaním, aké sú silné a slabé stránky jednotlivých technológií pre riešenie zadanej úlohy a následná vhodná aplikácia ich kombinácie pre efektívnejšie riešenie zložitých problémov a dosiahnutie čo najlepšieho výsledku.[229] **Hybridné systémy v širšom slova zmysle tak môžu znamenať posun k univerzálnejším, komplexnejším, odolnejším a sofistikovanejším systémom AI**.

V tiráži tejto publikácie je uvedené, že „žiadna časť tohto diela nie je vygenerovaná strojovým algoritmom“. Dovolíme si spraviť malú výnimku – uviesť popis hybridných systémov, s ktorým sa môžeme vo veľkej miere stotožniť:

Hybridný model uvažovania je v podstate navrhnutý tak, aby ponúkal komplexné riešenie prostredníctvom synergie rôznych paradigiem uvažovania, vďaka čomu je vhodný pre zložité scenáre, v ktorých by jeden prístup mohol byť nedostatočný.[230]

Prispôsobené (custom) generatívne modely

Vzhľadom na rozsiahle spektrum oblastí činnosti a typov úloh je vhodné uvažovať nad riešením prispôsobených (custom) LLMs, ktoré sú špecializované a osobitne vytrénované pre konkrétnu oblasť. **Ide o špecifické systémy genAI, ktoré môžu byť veľmi dobré v konkrétnej oblasti a v špecifickej množine poznatkov, ktoré využívajú**

---

hybrid-reasoning-model-is-now-available-in-amazon-bedrock/>

228 Jednu zo zaujímavých implementácií kombinácie viacerých typov pokročilého uvažovania uvádzame v časti *Agenti generatívnych systémov umelej inteligencie* v rámci popisu tzv. **multi-agent funkcionality a kooperatívneho uvažovania** modelu Grok-4 Heavy od xAI.

229 **Skutočnými hybridnými systémami** by sme mohli do určitej miery nazvať aj ostatné spôsoby vylepšovania genAI, ktoré v tejto kapitole uvádzame – používanie externých nástrojov, natívna kombinácia viacerých modalít, prispôsobené generatívne systémy alebo agenti AI.

230 Vygenerované 25. februára 2025 modelom DeepSeek-R1 (7.6B, Q4_K_M) v rámci lokálnej inštalácie (Alpaca na systéme Fedora Linux 41, 12th Gen Intel® Core™ i7-1260P × 16, 32 GiB RAM). Použitý lokálny model termín „hybrid reasoning model“ nepoznal – len na základe uvažovania dospel k solídnej analýze, k záveru ktorej patril i citovaný text.

pri generovaní odpovedí.

V súčasnosti sa vytváranie custom LLMs rieši prispôsobením na základe inštrukcií (napr. na čo je daný GPT zameraný, ako sa má správať, čoho sa má vyvarovať,...), špecifickým ladením a dotrénovaním v 2., resp. 3. kroku trénovania modelu a pridaním vlastného súboru poznatkov, na základe ktorého sa vykonáva tzv. retrieval-augmented generation (**RAG**).[231] Finálny výsledok sa využíva pri tvorbe odpovedí.

Prispôsobovanie modelu je do tej miery úspešné, do akej sme schopní pripraviť kvalitné súbory poznatkov pre prispôsobenie a primerané inštrukcie.[232]

Myšlienka custom LLMs bola implementovaná napr. v GPT-4 od OpenAI, v portfóliu prispôsobiteľných modelov napr. od Abacus.AI a vo veľkej miere v rôznych interných aplikáciách umelej inteligencie naprieč firemným sektorom, zdravotníctvom a pod.

Agenti generatívnych systémov umelej inteligencie (agentic AI)

Inteligentný agent je v oblasti AI vo všeobecnosti chápaný ako **činiteľ (systém, stroj), ktorý vníma prostredie, v ktorom pracuje, autonómne vykonáva akcie na dosahovanie cieľov a môže zlepšovať svoj výkon učením sa alebo získavaním znalostí**.[233]

---

231 Ide o techniku zameranú na zvýšenie presnosti a spoľahlivosti generatívnych modelov AI pomocou faktov získaných z externých zdrojov. Okrem natrénovaného modelu sa využíva i znalostná databáza dokumentov z konkrétnej oblasti. Pomocou jazykového modelu sa otázky z dokumentov i celý obsah dokumentov prevádza na sémantické vektory. Vektory z databázy, ktoré sú „najbližšie" vektoru otázky spolu s vektorom otázky sú následne vkladané do promptu a generuje sa odpoveď. Výsledkom sú presnejšie odpovede z danej oblasti, možnosť uvádzať zdroje, zapracovať modifikácie, napr. na vyhľadávanie v databáze dokumentov použiť vyhľadávač (search engine) a pod.
What Is Retrieval-Augmented Generation, aka RAG? [on-line]. [cit. 20. augusta 2024]. Dostupné na internete: <https://blogs.nvidia.com/blog/what-is-retrieval-augmented-generation/>
MONIGATTI, L. *Retrieval-Augmented Generation (RAG): From Theory to LangChain Implementation*. [on-line]. [cit. 17. januára 2024]. Dostupné na internete: <https://towardsdatascience.com/retrieval-augmented-generation-rag-from-theory-to-langchain-implementation-4e9bd5f6a4f2>

232 Dotrénovanie môže byť do istej miery rizikové, keďže môže nastať tzv. **katastrofická interferencia**. Ide o scenár, v ktorom model, ktorý už bol natrénovaný, intenzívne učíme niečo nové, pričom model sa dostane do stavu, že pôvodné znalosti stratí. Aby sa tomuto problému predišlo, s učením nových poznatkov musíme model učiť, resp. mu opakovať aj pôvodné znalosti.

233 Ide v zásade o akýkoľvek systém AI, ktorý pracuje v konkrétnom prostredí na základe vstupov z tohto

Základ tvorí vstup z prostredia (od jednoducho meraných dát až po dátový popis vnímaného „sveta“), výstup (od spracovaných dátových výsledkov až po akcie, ktoré v reálnom prostredí treba vykonať) a na cieľ zamerané jadro agenta (goal-directed agent, resp. racionálny agent).[234] **Ciele teda stanovujú ľudia, avšak najvhodnejšie postupy a činnosti, ktoré treba na dosiahnutie týchto cieľov vykonať, si samostatne vyberá inteligentný agent**.[235]

Tento princíp je aplikovaný i na agentov genAI – **ide o generatívne systémy, ktoré vykonávajú komplexné úlohy na pozadí konverzácie stroja s používateľom**. Prostredníctvom promptov je používateľ zadávateľom úloh (cieľov), pričom agent genAI sa postará o vykonanie komplexnej úlohy a v rámci konverzácie používateľovi poskytne výsledok, resp. vykoná konkrétnu akciu. Ak klasické systémy genAI sa zameriavajú na vytváranie obsahu na základe naučených vzorov, systémy Agentic AI rozširujú túto schopnosť uplatňovaním generatívnych výstupov na konkrétne ciele, a to aj vo veľmi dynamických prevádzkových prostrediach.[236]

---

prostredia, pričom na základe inteligentného spracovania vstupov v kontexte zadaných cieľov vykonáva výslednú činnosť v tomto prostredí. Môže ísť o čokoľvek od inteligentného termostatu, cez výrobnú linku na báze technológií AI alebo autonómne vozidlo, až po humanoidného robota atakujúceho ľudskú inteligenciu.

234 Na cieľ zameraný agent vykonáva akýkoľvek adaptívny a autonómny plán, ktorý ho privedie k maximalizácii tzv. účelovej funkcie. Napr. v prípade agenta využívajúceho učenie s posilnením sa pre dosahovanie požadovaného cieľa využíva funkciou odmeňovania (reward function), alebo v prípade využitia evolučného algoritmu hovoríme o funkcii vhodnosti (fitness function).
BRINGSJORD, S., GOVINDARAJULU, N. S. *Artificial Intelligence. The Stanford Encyclopedia of Philosophy (Summer 2020 Edition)*. [on-line]. [cit. 13. decembra 2024].
Dostupné na internete: <https://plato.stanford.edu/archives/sum2020/entries/artificial-intelligence/>

235 Ide o systémy AI známe aj pod názvom **Agentic AI**, vzhľadom na ich schopnosť konať nezávisle a zámerne, naviac s obmedzeným dohľadom. Komplexný systém Agentic AI môže pozostávať z jednotlivých samostatných agentov AI (napr. modelov strojového učenia), ktoré napodobňujú ľudský spôsob riešenia problémov.
STRYKER, C. *What is agentic AI?* [on-line]. [cit. 12. októbra 2025].
Dostupné na internete: <https://www.ibm.com/think/topics/agentic-ai>

236 Napríklad ChatGPT môže vygenerovať text, obrázok alebo kód, avšak agent genAI (agentic systém) môže tento vygenerovaný obsah použi na autonómne dokončenie zložitých úloh volaním externých nástrojov.
STRYKER, C. *What is agentic AI?* [on-line]. [cit. 12. októbra 2025].
Dostupné na internete: <https://www.ibm.com/think/topics/agentic-ai>

Princíp inteligentného agenta môže byť kombinovaný s používaním externých nástrojov generatívnymi modelmi, multimodalitou, využívaním komplexného rozhodovania a tvorbou prispôsobených modelov až do tej miery, že vzniká nepreberné množstvo oblastí nasadenia a možností využitia naprieč celým spektrom znalostnej spoločnosti.[237]

Roky 2025/2026 sú viacerými expertmi vnímané ako roky agentov AI, v ktorých – vzhľadom na vývoj a kvalitatívny pokrok tejto technológie – príde k ich širokému nasadeniu a uplatneniu.

Jednou z predpokladaných a očakávaných oblastí ich plošného nasadenia je oblasť informácií a webových služieb na internete. V zásade by agenti AI mali byť prostredníkmi medzi človekom a obsahom i službami dostupnými na internete. Agenti, vystupujúci v úlohe asistentov, by na základe ľudských požiadaviek vykonávali všetku činnosť (napr. objednanie taxi, rezerváciu letenky, vyhľadanie a spracovanie informácií k danej téme a pod.). Určite by išlo o neuveriteľné uľahčenie a zrýchlenie rôznych činností spojených s využívaním webových služieb. Na druhej strane – ak by bol tento koncept úspešný – internet by sa „ponoril do tmy“ strojového spracovania a stal by sa priestorom vyhradeným len pre stroje. **Človek by bol len konzumentom výsledkov bez akéhokoľvek hlbšieho kognitívneho zaangažovania a kritického posúdenia parciálnych činností spojených s vykonaním požadovanej úlohy**. Naviac, nástup komplexných asistentov je spájaný s prevádzkovateľmi tých najväčších a najlepších generatívnych systémov AI, ktorí majú technologické, výpočtové i dátové zázemie schopné realizovať tak komplexné systémy agentov AI. **Ich využitie ako prostredníkov medzi človekom a obsahom by znamenalo nesmiernu moc nad celým informačným tokom smerom k človeku a činnosťou realizujúcou ľudské zadania smerom k obsahu, technológiám a službám**. Moc, ktorá by bola sústredená v rukách niekoľkých technologických gigantov.[238]

---

237 Kombinácia môže byť nielen statická, ale aj dynamická – vytváraná priamo v rámci riešenia zadanej úlohy. Napríklad DeepAgent od Abacus.AI na automatizovanie zložitejších úloh dynamicky vytvára potrebných agentov za behu (on-the-fly agents).
DeepAgent Automatically Creates On-The-Fly Agents To Automate More Complex Tasks. [on-line]. [cit. 12. októbra 2025].
Dostupné na internete: <https://x.com/abacusai/status/1954272264066912479>

238 BARR, A. *AI agents could make the internet go dark*. [on-line]. [cit. 24. marca 2025].
Dostupné na internete: <https://www.businessinsider.com/ai-agents-internet-dark-google-openai-anthropic-2025-1>

I tento príklad predpokladaného využitia svedčí o tom, ako veľmi **v oblasti špičkových technológií AI je a bude potrebná diverzita a otvorenosť**.[239]

Oblasť agentov AI je dynamicky sa rozvíjajúcou oblasťou nasadenia moderných systémov genAI. Riešenia ponúkajú nielen zavedení veľkí hráči (napr. Microsoft so špecializovaným Copilotom pre pracovníkov kybernetickej bezpečnosti s integrovanými agentmi AI realizujúcimi špecifické úlohy ochrany, detekcie a prevencie)[240], ale i firmy špecializujúce sa priamo na tvorbu agentov AI, resp. celých ekosystémov vytvárania agentov AI zákazníkmi na mieru.[241]

9. júla 2025 bol predstavený nový model Grok od spoločnosti xAI vo verziách Grok-4 a Grok-4 Heavy s pokročilým uvažovaním, špecializovanými funkciami kódovania, textovou i ohlásenou vizuálnou modalitou a vynikajúcim výkonom v mnohých pokročilých testoch AI.

Grok-4, osobitne vo verzii Heavy, je i tzv. **multi-agent systémom so schopnosťou kolaboratívneho uvažovania**. Ide o novú funkcionalitu genAI systému, ktorý pre riešenie úloh vytvára viacero agentov pracujúcich paralelne. Títo agenti nezávisle riešia problémy, porovnávajú výsledky a kombinujú poznatky, aby poskytli čo najlepšiu odpoveď a zvládali zložité úlohy využívaním kolaboratívneho (tímového) uvažovania.[242]

---

239 Tejto problematike sa aspoň okrajovo venujeme v časti *Destilácia znalostí (knowledge distillation) a vytváranie malých lokálnych modelov* v rámci tejto kapitoly.

240 JAKKAL, V. *Microsoft unveils Microsoft Security Copilot agents and new protections for AI*. [on-line]. [cit. 26. marca 2025].
Dostupné na internete: <https://www.microsoft.com/en-us/security/blog/2025/03/24/microsoft-unveils-microsoft-security-copilot-agents-and-new-protections-for-ai/>

241 Jednou z najlepších a dynamicky rozvíjajúcich sa firiem vývoja a tvorby na mieru šitých agentov a asistentov AI je Abacus.AI.
An AI Super Assistant For Enterprises And Professionals. [on-line]. [cit. 27. marca 2025].
Dostupné na internete: <https://abacus.ai/>

242 V multiagentnom nastavení riešia agenti zadanie súbežne, ako študijná skupina. Spolupracujú tak, že si rozdelia úlohy, pracujú nezávisle a potom zdieľajú svoje zistenia, porovnávajú svoje riešenia, identifikujú nezrovnalosti a kombinujú poznatky, aby vytvorili a spoločne vylepšili konečnú odpoveď. Tento postup tvorcovia považujú za vhodný pre zložité problémy, pretože zvyšuje presnosť prostredníctvom krížovej kontroly a tímového uvažovania. Do určitej miery ide o napodobňovanie ľudských skupín pracujúcich na zadanom probléme.
Systém je postavený na integrácii natívnych nástrojov, takže agenti môžu čerpať údaje v reálnom čase alebo používať kódovacie nástroje s obrovským kontextovým oknom (256 000 tokenov) pre hlbšie

Multi-agent funkcionalita so schopnosťou kolaboratívneho, resp. tímového uvažovania korešponduje s naším pohľadom na tzv. skutočné hybridné systémy, o ktorých sme písali v časti Hybridné modely uvažovania.

S možnosťami a potenciálom agentov AI sa prienik umelej inteligencie do reálneho sveta stáva emergentným, mnohokrát môže pôsobiť disruptívne a javí sa perzistentným. Princíp inteligentného agenta AI teoreticky vytvára priestor pre rast k skutočnej umelej inteligencii konajúcej v reálnom svete ako samostatná entita.[243]

Zväčšovanie kontextového okna

Kontextové okno (context window) v rámci veľkých jazykových modelov (LLM) označuje počet tokenov, s ktorými LLM pracuje pri generovaní odpovede, resp. ktoré model dokáže mať na vstupe pri generovaní výstupu. Kontextové okno tak predstavuje množstvo textu, ktoré môže jazykový model spracovať naraz počas konverzácie alebo úlohy.[244]

---

analýzy. Autori popracovali na výkone, ktorý sa v čase testovania zvýšil desaťnásobne (predpokladáme že voči Grok 3).
Introducing Grok 4, the world's most powerful AI model. [on-line]. [cit. 11. júla 2025].
Dostupné na internete: <https://x.com/xai/status/1943158495588815072>

243 Systémy agentic AI tvoria určitý posun k filozofii AGI v tom zmysle, že celý komplexný proces úloh sa vykonáva samostatne a spúšťa autonómne procesy, ktoré sú naviac schopné vykonávať činnosť aj v externom prostredí.
To je niečo, čo treba mať pod ľudskou kontrolou, inak ťažko zabezpečiť ich dôveryhodnú činnosť.

244 Napr. kontextové okno pri GPT-3 má veľkosť 2000 tokenov, pri GPT-3.5 od 8000 do cca. 16000, no pri GPT-4 je to už od 32000 do 128000. GPT-4o a model o1 majú tiež 128000 tokenov.
Model Gemini 2.0 však pracuje s 1-2 miliónmi tokenov a model Claude 3.5 používa v rámci kontextového okna 200000 tokenov. Grok-4 / Heavy má kontextové okno s 256 000 tokenmi **I na počte tokenov kontextového okna je vidieť rozdiely vo vývoji modelov najväčších hráčov v oblasti genAI** (OpenAI, Google, Anthropic, xAI,...).
Context Window for LLMs. [on-line]. [cit. 20. novembra 2024].
Dostupné na internete: <https://www.hopsworks.ai/dictionary/context-window-for-llms>
OpenAI Platform - Models. [on-line]. [cit. 5. decembra 2024].
Dostupné na internete: <https://platform.openai.com/docs/models/gp>
LUKEŠ, D. *Long Context Windows: Opportunities and Challenges*. [on-line]. [cit. 10. decembra 2024].
Dostupné na internete: <https://ai-trends.notion.site/Long-Context-Windows-Opportunities-and-Challenges-1404869badd7804f87b9f596fdb1fee6>
Introducing Grok 4, the world's most powerful AI model. [on-line]. [cit. 11. júla 2025].
Dostupné na internete: <https://x.com/xai/status/1943158495588815072>

Kontextové okno teda vyjadruje, koľko informácií môžeme poskytnúť výzve (promptu), aby sme získali požadovaný výstup. Čím väčšie je kontextové okno, tým viac informácií na vstupe dokáže genAI spracovať pri generovaní výstupu. Od veľkosti kontextového okna tiež závisí, do akej miery vie systém generovať odpovede relevantné v rámci celého obsahu komunikácie.[245]

Veľkosti kontextového okna jednotlivých LLMs, ktoré ich tvorcovia uvádzajú, sa viažu na anglický text. Pri iných jazykoch býva veľkosť kontextového okna spravidla iná, väčšia. Iné jazyky zväčša vyžadujú pre analogickú činnosť o 50-200% viac tokenov. No sú i jazyky (napr. pandžábčina), ktoré vyžadujú tokenov desaťnásobne viac.[246]

Zväčšovaním kontextového okna môžu súčasné veľké jazykové modely nielen spracúvať oveľa väčšie množstvo dát na vstupe, ale – a to predovšetkým – zvyšovať kvalitu svojich výstupov. Podobne ako pri škálovaní generatívnych systémov AI prostredníctvom zvyšovania počtu neurónov a množstva trénovacích dát, aj zväčšovaním kontextového okna môžeme získať „viac inteligencie“. V tomto prípade však nejde o získavanie inteligencie „hrubou silou“, ale skôr o snahu lepšie porozumieť vstupu (promptu) v kontexte celej komunikácie a následne generovať relevantnejšie odpovede.

Cieľom v tejto oblasti je tzv. *infinite context window*, t.j. vstup bez obmedzení, a uchopenie celej konverzácie, čo by podľa viacerých expertov mohlo znamenať nielen kvalitatívny posun vo vývoji súčasných generatívnych systémov, ale i pokrok na ceste k AGI.[247]

Potreba kontextového okna poukazuje na neschopnosť veľkých jazykových modelov „čítať“ a reagovať ľudským spôsobom myslenia. Vzhľadom na mechanizmus pozornosti (self-attention a multi-head attention) je spracovanie

245 Väčšia veľkosť kontextového okna zvyšuje schopnosť systému vykonávať kontextové učenie sa v rámci promptu. Táto schopnosť umožňuje LLM vytvárať kontextuálnejšie relevantné odpovede využitím komplexnejšieho pochopenia vstupných údajov.
Context Window for LLMs. [on-line]. [cit. 20. novembra 2024].
Dostupné na internete: <https://www.hopsworks.ai/dictionary/context-window-for-llms>

246 LUKEŠ, D. *Long Context Windows: Opportunities and Challenges*. [on-line]. [cit. 10. decembra 2024].
Dostupné na internete: <https://ai-trends.notion.site/Long-Context-Windows-Opportunities-and-Challenges-1404869badd7804f87b9f596fdb1fee6>

247 Dr. Eric Schmidt, Chair of the Special Competitive Studies Project, speaks at SCSP's inaugural AI+Energy Summit.
Fireside - Eric Schmidt. [on-line]. SCSP, 2024, 2:42. [cit. 20. novembra 2024].
Dostupné na internete: <https://www.youtube.com/watch?v=oC46CzxT750>

kontextových okien výpočtovo veľmi náročné a zväčšovanie kontextového okna si pri súčasných technológiách vyžaduje vždy vyššie nároky na výpočtové zdroje (procesorový čas, pamäť,...).

Destilácia znalostí (knowledge distillation) a vytváranie malých lokálnych modelov

Schopnosti veľkých jazykových modelov a vo všeobecnosti pokročilých generatívnych systémov umelej inteligencie sú limitované dostatočne výkonnými výpočtovými prostriedkami a ich energetickým apetítom, prípadne dostupnosťou týchto „veľkých“ systémov kdekoľvek a kedykoľvek.

Pomerne dôležitým smerov vývoja je snaha preniesť schopnosti týchto veľkých modelov do menších systémov, ktoré by mohli fungovať lokálne, t.j. nezávisle od on-line prepojenia na veľké dátové centrá a na bežnom technickom vybavení (napr. v mobile, prenosnom počítači, inteligentnej kamere, priemyselnom zariadení vo výrobnej hale a pod.).

Jednou z praktických realizácií týchto snáh je i tzv. destilácia znalostí (knowledge distillation). Ide o **proces prenosu kľúčových poznatkov a schopností z veľkých špičkových modelov do menších a efektívnejších modelov**.

V zásade ide o trénovanie kompaktnejšieho modelu, ktorý by napodobňoval väčší a zložitejší model, pričom na rozdiel od veľkého modelu sa nevykonáva komplexné trénovanie, ktoré sme uvádzali v predchádzajúcej kapitole, ale kompaktný model sa trénuje tak, aby sa jeho výsledky zhodovali s predikciami veľkého modelu. Takto vytrénovaný kompaktný model môže byť aj tisíckrát menší a rýchlejší, resp. pracujúci nie na superpočítačoch, resp. veľkých cloudových platformách, ale na modernom mobilnom zariadení, ktoré bežne nosíme vo vrecku.[248]

V súčasnosti, keď môžeme hovoriť o tých najlepších uzavretých (proprietárnych) modeloch ako o supermodeloch AI,[249] sa destilácia znalostí zároveň javí ako účinný prostriedok prenosu pokročilých schopností z týchto proprietárnych modelov do menších a prístupnejších modelov s otvoreným zdrojovým kódom. Pre budúci vývoj a dopad technológií umelej inteligencie na spoločnosť, ktorá je nielen znalostná, ale i slobodná a na hodnotách postavená, považujeme divergenciu

248 Tieto kompaktné lokálne modely majú v súčasnosti 1-100 miliárd parametrov a sú schopné pracovať na bežných zariadeniach so základnou hardvérovou podporou pre operácie AI.

249 Hovoríme o supermodeloch z pohľadu technologickej vyspelosti a náročnosti, nie z pohľadu schopností AGI.

systémov AI a ich schopností za veľmi potrebnú a prínosnú.

Čínsky model DeepSeek-R1, ktorý je výkonovo na úrovni špičkových modelov OpenAI o1, Gemini 2.0 a Claude 3.5, sme na začiatku kapitoly už spomínali. Vzhľadom na to, že tento veľmi kvalitný model je vydaný pod open-source licenciou,[250] stáva sa predmetom nielen lokálnych inštalácií a prevádzky, ale aj širšieho skúmania a prepojenia s inými technológiami. V oblasti destilácie znalostí je tento model využitý na doladenie menších, resp. špecifických modelov. Výsledné modely tak v sebe kombinujú znalosti z pôvodných modelov (napr. Qwen, Llama,...) s pokročilými vzorcami uvažovania DeepSeek-R1.[251] **Destilácia znalostí v kombinácii so schopnosťou uvažovať kvalitatívne i výkonovo posúva malé modely, ktoré je možné prevádzkovať i lokálne, na oveľa vyššiu úroveň systémov umelej inteligencie**.

Uvoľnenie čínskeho modelu DeepSeek-R1 v januári 2025 znamenalo nielen veľkú výzvu pre najlepšie modely genAI, ale vzhľadom na veľkú publicitu predstavovalo aj enormný záujem používateľov z celého sveta. Ani nie týždeň po uvoľnení modelu pre verejnosť boli zistené závažné nedostatky v bezpečnosti modelu i v ochrane dát a osobných údajov.[252] Na rovinu treba uviesť, že i keď v oveľa menšej miere, predsa však podobné nedostatky majú aj iné modely, ktoré sú prevádzkované cez internet, resp. integrované do operačných systémov počítačov a mobilných zariadení. **Odpoveďou na tento problém môžu byť lokálne modely, ktoré vďaka destilácii znalostí, kombinácii so schopnosťou uvažovať, resp. implementovať vzorce uvažovania a open-source licenciou môžu**

---

250 Existujúce open-source systémy AI bývajú vydané pod niektorou z open-source licencií, napr. MIT, Apache a pod.
OSI Approved Licenses. [on-line]. [cit. 20. februára 2025].
Dostupné na internete: <https://opensource.org/licenses>
Open Source Iniciatíva (OSI) pripravila **definíciu AI s otvoreným zdrojovým kódom**, t.j. základný rámec, ktorý musí systém AI spĺňať, aby mohol byť vydaný pod niektorou z licencií open-source.
The Open Source AI Definition – 1.0. [on-line]. [cit. 9. marca 2025].
Dostupné na internete: <https://opensource.org/ai/open-source-ai-definition>

251 *Ollama: deepseek-r1*. [on-line]. [cit. 12. februára 2025].
Dostupné na internete: <https://ollama.com/library/deepseek-r1>

252 Základné problémy (nie všetky, postupne boli zistené i ďalšie) sú uvedené vo varovaní Národného bezpečnostného úradu.
Upozornenie na AI model. [on-line]. [cit. 2. februára 2025].
Dostupné na internete: <https://www.nbu.gov.sk/upozornenie-na-ai-model/>

byť pomerne výkonnou a dôveryhodnou alternatívou známych systémov genAI.[253]

Generatívne systémy AI ako operačné systémy[254]

Jedným zo špekulatívnych smerov vývoja generatívnych systémov AI je využitie tejto technológie ako technologického základu emergentného operačného systému (OS), resp. technologicky prepojeného ekosystému.[255] **Išlo by teda nie o jednotlivé nástroje a systémy umelej inteligencie, ani nie o sofistikované systémy a riešenia, ktoré v tejto kapitole uvádzame, ale o prepojený celok, v rámci ktorého rôznorodé technológie AI nielenže spolupracujú na riešení zadaného problému, ale tento problém aj integrálne riešia**.[256]

Na schematickom obrázku je vyjadrená analógia genAI/LLM OS so súčasnými operačnými systémami: CPU, RAM (~ contex window), disk, softvérové nástroje, počítačová sieť,

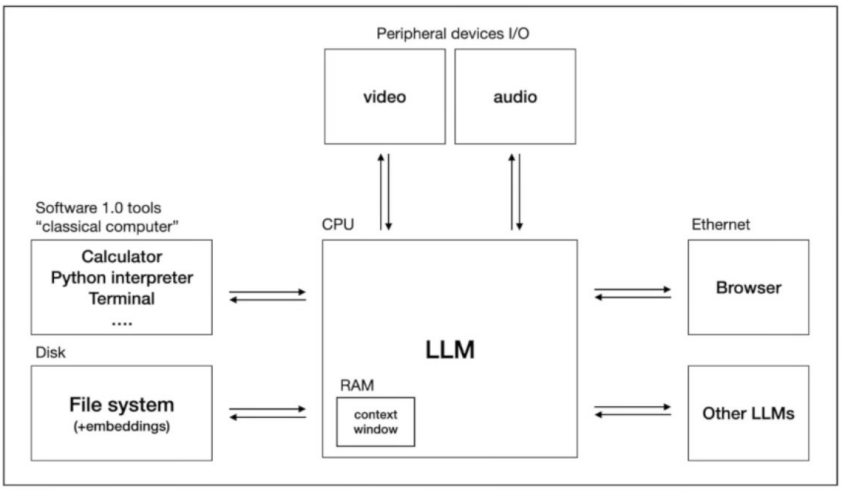
253 Záujemcom o lokálnu prevádzku genAI systémov môžeme odporučiť open-source systém *ollama*, ktorý zabezpečuje beh veľkých jazykových modelov (DeepSeek, Llama, Mistral, Phi-4, Gemma a niekoľko desiatok ďalších). Ollama síce neobsahuje grafické rozhranie, avšak je ho možné kombinovať s niektorou z existujúcich grafických nadstavieb (autor na systéme Linux využíva grafické rozhranie Alpaca, ktoré je dostupné i prostredníctvo Flathubu).
Podrobnejšie informácie o možnostiach a spôsoboch optimalizácie veľkých genAI modelov i dôvodoch ich prevádzky na lokálnych zariadeniach sú uvedené v prílohe *Lokálna prevádzka generatívnych systémov AI*.
Ollama: deepseek-r1. [on-line]. [cit. 12. februára 2025].
Dostupné na internete: <https://ollama.com/library/deepseek-r1>
GitHub – Alpaca. [on-line]. [cit. 12. februára 2025].
Dostupné na internete: <https://github.com/Jeffser/Alpaca>

254 Ide o myšlienku Andreja Karpathyho, slovensko-kanadského počítačového vedca, ktorý vytvoril prvý kurz deep learningu na Stanfordskej univerzite, bol spoluzakladateľom OpenAI a viedol tím vývoja umelej inteligencie a autopilota v Tesle.
KARPATHY, A. *Intro to Large Language Models*. [on-line]. [cit. 13. februára 2024].
Dostupné na internete: <https://youtu.be/zjkBMFhNj_g>

255 Operačný systém (OS) je základný softvér, ktorý zabezpečuje činnosť počítača, umožňuje interakciu medzi počítačovým hardvérom a užívateľskými aplikáciami. OS spravuje zdroje počítača a poskytuje programom rozhranie na prístup k týmto zdrojom. Operačný systém tiež spracúva systémové dáta a vstupy od používateľa a odpovedá alokovaním interných zdrojov a spravovaním úloh počítača ako služby pre používateľa.
Operating system. [on-line]. [cit. 5. decembra 2024].
Dostupné na internete: <https://en.wikipedia.org/wiki/Operating_system>

256 V zásade by išlo o vylepšenie a synergiu viacerých prístupov, ktoré sme doteraz uvádzali.

periférne zariadenia, multiprocesorové spracovanie úloh, špekulatívne vykonávanie inštrukcií,... jednoducho softvér a hardvér, ktorý sa ako celok podieľa na riešení zadaných úloh.



Obr. č. 10: LLM OS – genAI ako operačný systém.[257]

Inú analógiu technologicky prepojeného ekosystému ponúka porovnanie s operačnými systémami s otvoreným kódom (open-source), ktoré sú technologicky prepojené, ich kód je zdieľaný, rozvíjajú sa, čerpajúc jeden od druhého, vyvíjajú sa nové technológie na základe predchádzajúcich a pod. Ide o spôsob vývoja, ktorý je veľmi dynamický a rôznorodý, nie nepodobný viacerým obdobiam hľadania, bádania a vývoja systémov umelej inteligencie.[258]

V kontexte neustáleho vylepšovania generatívnych systémov, ktoré sme doteraz načrtli, by integrácia, kooperácia a orchestrácia rôznorodých technológii AI mohla predstavovať ďalší evolučný posun k skutočnej umelej inteligencii. Domnievame sa však, že bez skutočných koncepčných prelomov, o ktorých budeme uvažovať v 5. kapitole, nie sme schopní vytvoriť, resp. spraviť skutočne revolučný krok na ceste k AGI (minimálne na úrovni HLAGI, t.j. skutočnej umelej inteligencie úrovne človeka).

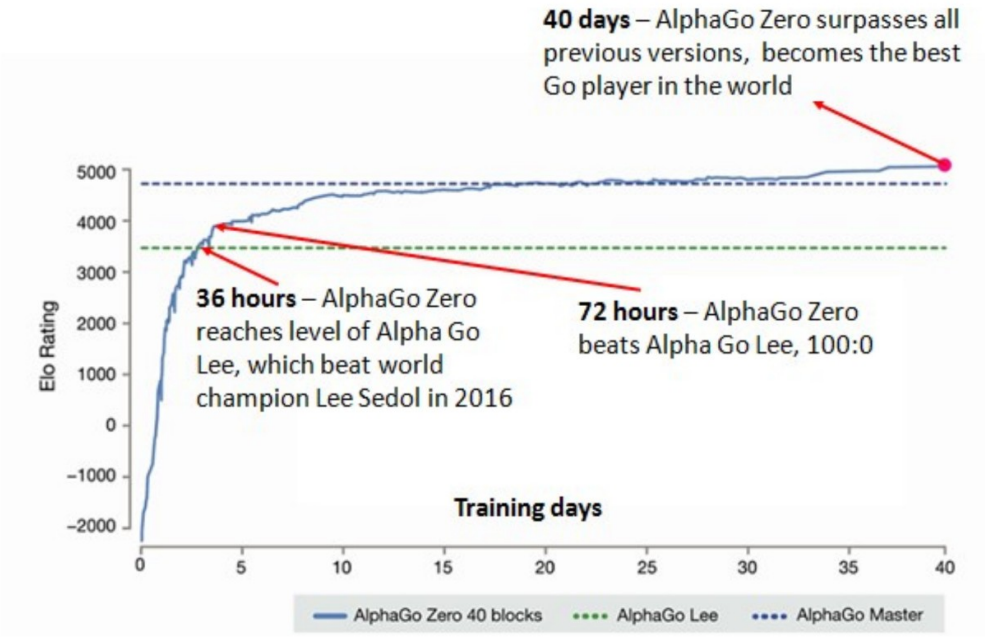
257 KARPATHY, A. *Intro to Large Language Models*. [on-line]. [cit. 13. februára 2024]. Dostupné na internete: <https://youtu.be/zjkBMFhNj_g>

258 Túto problematiku sme diskutovali v kapitole 1.2: Neexistencia presnej definície AI a viac menej ignorancia základných kategorizácií sú živnou pôdou pre veľmi rôznorodý a kreatívny rozvoj tejto oblasti, keďže bádateľov vedie len hrubý zmysel pre toto smerovanie a snaha etablovať sa, či dostať sa do popredia v tejto emergentnej oblasti ... rovnako búrlivý je i rozvoj technológií a príbuzných vedných odborov, čo sa v nemalej miere podieľa na tvorbe nových riešení umelej inteligencie.

Sebazdokonaľovanie podľa vzoru AlphaGo[259]

Technika sebazdokonaľovania (self-improvement) technológií AI je známa napríklad z realizácie systému AlphaGo od Google, ktorý v hre Go zvíťazil nad najlepšími hráčmi sveta.[260]

AlphaGo (a jeho nástupcovia až po Alpha Zero) v sebe kombinuje hlboké neurónové siete s pokročilými vyhľadávacími algoritmami. V prvej fáze bol systém AlphaGo zapojený v mnohých amatérskych hrách Go, aby sa mohol naučiť, ako ľudia hrajú túto hru.



Obr. č. 11: Fázy učenia AlphaGo Zero.[261]

V druhej fáze nasledovalo učenie sebazdokonaľovaním – AlphaGo niekoľko tisíckrát hral proti rôznym verziám seba samého, pričom sa zakaždým učil zo svojich chýb metódou

259 KARPATHY, A. *Intro to Large Language Models*. [on-line]. [cit. 13. februára 2024]. Dostupné na internete: <https://youtu.be/zjkBMFhNj_g>

260 **Víťazstvo AlphaGo inšpirovalo novú éru systémov AI. Bolo presvedčivým dôkazom, že základné neurónové siete sa dajú aplikovať na zložité oblasti, zatiaľ čo použitie učenia s posilňovaním ukázalo, ako sa stroje môžu naučiť riešiť neuveriteľne ťažké problémy samé, jednoducho metódou pokus-omyl**.
Schopnosť AlphaGo pozerať sa dopredu a plánovať sa stále využíva aj v dnešných systémoch AI. *AlphaGo*. [on-line]. [cit. 19. novembra 2024]. Dostupné na internete: <https://deepmind.google/research/breakthroughs/alphago/>

261 *Deep Reinforcement Learning for Natural Language Processing*. [on-line]. [cit. 20. augusta 2024]. Dostupné na internete: <https://sites.cs.ucsb.edu/~william/papers/ACL2018DRL4NLP.pdf>

učenia s posilnením. Postupom času sa AlphaGo neustále zlepšoval a stával sa lepším hráčom, prekonávajúc tak schopnosti najlepších hráčov sveta (obrázok č. 11).[262]

Pri AlphaGo je učenie sebazdokonaľovaním pomerne jednoduché – veľa realizovaných hier v sandboxoch (virtuálnych oddelených prostrediach), jednoduché a lacné vyhodnocovanie funkcie odmeňovania (reward function) a automatizovanie celého procesu. Zdokonaľovanie prebieha na miliónoch realizovaných hier s jediným kritériom, ktorým je pravdepodobnosť výhry. Systém takto nie je konfrontovaný s imitáciou najlepších hráčov, ide za možnosti človeka.

Porovnávajúc s AlphaGo/AlphaZero, pri generatívnych modeloch AI v súčasnosti robíme len trénovanie prvej fázy, t.j. učenie sa z dát sveta, ľudské značkovanie, ľudské, resp. imitované ľudské učenie s odmeňovaním. Nie sme schopní ísť za hranicu možností tohto spôsobu učenia a z toho prameniacej presnosti (v rámci algoritmov genAI).

Vnímajúc úspechy druhej fázy trénovania AlphaGo sa však môžeme zamýšľať nad jej analógiou pre oblasť veľkých jazykových modelov a genAI. Základným problémom sú chýbajúce všeobecné kritériá odmeňovania (reward function). Neriešime odmeňovanie v rámci pravidiel nejakej hry, ale v priestore jazyka a širokého poľa typov riešených úloh. Chýba nám nielen jednoducho vyhodnotiteľná, ale v zásade akákoľvek relevantná všeobecná funkcia odmeňovania.

V zásade sme schopní dosiahnuť, resp. vytvoriť funkciu odmeňovania v úzkej oblasti činnosti genAI, no je veľkým problémom to realizovať vo všeobecnej rovine.

Naviac, pri generatívnych systémoch AI by mohol byť vo všeobecnosti problém nielen s hľadaním a implementáciou správnej funkcie odmeňovania, ale aj s postupmi, ktoré k „odmene“ vedú. Totižto implementácia pravidiel (napr. etických) v generatívnych systémoch AI je oveľa náročnejšia, než implementácia pravidiel hry Go v rámci AlphaGo.

Ďalším veľkým problémom realizácie druhej fázy trénovania genAI sú dáta. V súčasnosti sa pomaly blížime k obsiahnutiu všetkých dostupných on-line dát pre klasické trénovanie veľkých generatívnych modelov. Vedecká obec sa viacmenej zhoduje v názore, že

262 Jedinou funkciou odmeňovania (reward) je výhra v hre. V zásade AIphaGo hľadá čokoľvek, čo vedie k výhre (naplneniu reward funkcie), aby sa dokázal zlepšovať (samozrejme v rámci implementovaných pravidiel hry Go).
AlphaGo. [on-line]. [cit. 19. novembra 2024].
Dostupné na internete: <https://deepmind.google/research/breakthroughs/alphago/>

pre kvalitatívny pokrok v príprave lepších modelov genAI bude treba nachádzať ďalšie zdroje kvalitných dát, nakoľko syntetické trénovacie dáta – tak, ako ich vieme pripraviť v súčasnosti – vedú skôr k zníženiu, a nie k zvýšeniu kvality pripravovaných modelov.[263]

Pokroky v neurovede ako základ pre nové modely

V našej rozprave o symbolických a subsymbolických systémoch AI sme na margo mnohých subsymbolických technológií uvádzali, že ich vývoj bol inšpirovaný pokrokom v neurovede, pričom sa tieto systémy usilujú o emuláciu činnosti mozgu na úrovni neurónov (napr. konvolučné neurónové siete ConvNets/CNN, ktoré boli navrhnuté podľa dobovej znalosti činnosti vizuálneho cortexu v mozgu). **V symbióze aktuálneho vývoja v neurovede a dosiahnutých úspechov v oblasti (nielen) generatívnych systémov AI sme svedkami ďalšieho posúvania schopností pokročilých systémov AI**.

Príkladom môže byť v decembri 2024 predstavený model MovieNet, ktorý sa snaží spracúvať pohyblivé obrazy podobne, ako to robí ľudský mozog. Ide v zásade o simuláciu schopnosti mozgu vytvárať ucelené vizuálne príbehy z meniacich sa scén a rozpoznávať prirodzené filmové scény (simuláciu spôsobu, akým neuróny spracovávajú dynamické vizuálne podnety a využitie stratégií kódovania priestorovo-časových transformácií v mozgu).[264]

Benefitom uvedeného prístupu je nielen pokrok v spôsobe, rýchlosti a presnosti spracovania (klasifikácii) vizuálnych pohyblivých vnemov, ale i nízka réžia a ekologická

263 Viac o tejto problematike pojednávame v závere kapitoly.

264 Prirodzeným vizuálnym scénam dominuje časopriestorová dynamika obrazu, ale nie je jasné, ako vizuálny systém integruje „filmové“ informácie v čase. Pri vývoji MovieNet boli charakterizované recepčné polia neurónov zrakového nervu pomocou podnetov s riedkym šumom a analýzy spätnej korelácie. Neuróny rozpoznávali filmy s trvaním ~ 200 - 600 ms s definovanými začiatočnými a koncovými podnetmi. Trvanie filmov od začiatku po koniec reakcie bolo vyladené na základe zmyslových skúseností prostredníctvom hierarchického algoritmu. Neuróny kódovali rodiny obrazových sekvencií podľa trigonometrických funkcií. Sekvencia hrotov (spike) a tok informácií naznačujú, že opakujúce sa motívy obvodov sú základom detekcie filmov. Princípy topografickej plasticity žabej sietnice a kortikálnych jednoduchých buniek sa využívajú v sieťach strojového učenia na rozpoznávanie statických obrazov, čo naznačuje, že objavy princípov kódovania filmov v mozgu, napríklad spôsob kódovania obrazových sekvencií a trvania, môžu byť prínosom pre technológiu rozpoznávania filmov. HIRAMOTO, M., CLINE, T. H. *Identification of movie encoding neurons enables movie recognition AI*. [on-line]. [cit. 5. decembra 2024].
Dostupné na internete: <https://www.pnas.org/doi/10.1073/pnas.2412260121>

šetrnosť (menšie požiadavky na množstvo trénovacích dát a rozsah trénovania nakoľko sa znížil objem údajov a počet krokov na dokončenie klasifikačnej úlohy).

Primárne využitie je v oblasti detekcie anomálií s využitím v medicíne (napr. včasné rozpoznávanie niektorých závažných chorôb), vo farmaceutickom priemysle (dopad na činnosť buniek a ich reakcie), v oblasti autonómneho riadenia (rozlišovanie a rýchla reakcia na zmeny v prostredí), v priemysle, v armádnych a bezpečnostných systémoch a pod.

Na základe postrehov, ktoré sme v závere kapitoly 1.7. uviedli na margo rozdielov medzi biologickými a umelými neurónovými sieťami i v kontexte perspektívneho rozvoja neuromorfných systémov a celkového využívania poznatkov neurovied, predpokladáme veľký potenciál ďalšieho vývoja nielen v oblasti modelov genAI, ale predovšetkým v hľadaní nových algoritmov a prístupov, ktoré nás pri vývoji umelej inteligencie posunú ďalej.

Zmeny v spôsobe využívania nástrojov umelej inteligencie[265]

Prelom, ktorý znamenalo sprístupnenie ChatGPT spoločnosťou OpenAI a následný úspech generatívnych systémov ide ruka v ruke so zmenami v spôsobe využívania systémov AI, inováciami a novými možnosťami v oblasti interakcie medzi človekom a strojovou inteligenciou.

Základom pre úspech genAI, osobitne veľkých jazykových modelov, je konverzačné rozhranie umožňujúce komunikáciu v bežnom jazyku formou dialógu a pokročilej – akoby tímovej – spolupráce medzi človekom a strojom. Behom takmer troch rokov, ktoré uplynuli od zverejnenia ChatGPT, sa v tejto oblasti vykonalo veľmi veľa. Napríklad:

- uchovávanie histórie konverzácií a možnosti sémantického vyhľadávania v komunikácii;
- implementácia interaktívnych funkcií, napr. navigácie vo verziách, možnosti úprav,...
- zakomponovanie rozšírených nástrojov, napr. generovanie kódu, pokročilá analýza údajov, uvažujúce modely, ...

265 LUKEŠ, D. *Trend 2: Interface Innovations: (Re)defining the user experience of AI*. [on-line]. [cit. 10. decembra 2024].
Dostupné na internete: <https://ai-trends.notion.site/Trend-2-Interface-Innovations-Re-defining-the-user-experience-of-AI-13a4869badd7804e87fcd6bdc371c156>

- integrácia multimodálnych schopností, napr. hlasové režimy, generovanie obrázkov, práca so súbormi a vloženými štruktúrovanými i neštruktúrovanými dátami a pod.

V rámci rôznych modelov a technológií sa rozvinuli o rôzne prístupy k riešeniu spôsobu využívania nástrojov genAI, napríklad:

- schopnosť pracovať s už vygenerovaným obsahom (ChatGPT Canvas, Claude Artifacts, ...);
- zameranie sa na workflow, pracovné postupy a životný cyklus dokumentov (Google NotebookLM, ...);
- integrácia technológií AI do celého softvérového ekosystému (Microsoft Copilot, ...);
- špecializované nástroje zamerané napr. na vyhľadávanie (Perplexity), písanie (Notion AI), výskum (Elicit);

I napriek veľkému pokroku v intuitívnosti a spoľahlivosti konverzačných rozhraní existujú viaceré aktuálne výzvy a úlohy, ktoré bude treba v najbližšom období riešiť, inak bude využiteľnosť týchto systémov stagnovať. Príkladom:

- treba riešiť aktuálne technologické, bezpečnostné, psychologické a sociologické problémy, ktorým sa venujeme v 2. kapitole;
- generatívne systémy sa vzhľadom na svoje možnosti a implementované funkcie stávajú pre bežného používateľa zložitými: používatelia zostávajú pri veľmi jednoduchom využívaní týchto nástrojov a často nepoznajú všetky možnosti;
- je treba zapracovať na možnostiach využiteľnosti systémov pre tímovú spoluprácu viacerých požívateľov, na organizácii a správe používateľov v tímoch i na lepších možnostiach dynamického narábania s vygenerovanými výsledkami;
- integrácia genAI s rôznymi inými technológiami, robotickými počnúc a končiac pri tradičných rozhraniach a technológiách;
- komplexná integrácia na úrovni operačných systémov a technologických zariadení pri zachovaní vysokých štandardov ochrany súkromia, spoľahlivosti a bezpečnosti;
- návrh a implementácia regulácií, ktoré nebránia inovácii a použiteľnosti, no chránia pred zneužitím a rizikami systémov AI.

Využívanie nástrojov umelej inteligencie s nástupom generatívnych systémov

smeruje k prirodzenému ľudskému spôsobu komunikácie a k aplikácii postupov, ktoré majú potenciál radikálne meniť obsah i spôsob našej práce. Dôsledky tohto smerovania majú vplyv nielen na rozvoj znalostnej spoločnosti, ale – a to predovšetkým – na ľudskú psychiku a sociologický rozmer fungovania spoločnosti.

Testovanie vlastností generatívnych systémov AI

Diskutovať o vylepšovaní a schopnostiach genAI nie je možné bez reálnej a efektívnej možnosti porovnávať, hodnotiť a testovať vlastnosti jednotlivých modelov a modelových množín.

V súčasnosti existuje pomerne veľa testov na vyhodnocovanie a porovnávanie schopností systémov generatívnej umelej inteligencie, z ktorých niektoré môžeme v skratke opísať:[266]

- MMLU – Massive Multitask Language Understanding je referenčný test určený na hodnotenie všeobecných znalostí a porozumenia modelu v širokom rozsahu tém. Patria sem predmety ako história, prírodné vedy, literatúra a právo. Skóre MMLU odráža, ako dobre si model poradí s rôznymi otázkami a informáciami, a tak poskytuje meradlo miery jazykového porozumenia a schopnosti riešiť problémy.
- HumanEval – referenčné kritérium, ktoré sa používa na hodnotenie schopností modelu programovať a vytvárať softvérový kód. Obsahuje súbor programovacích úloh zameraných na hodnotenie toho, ako efektívne dokáže model písať funkčný kód, riešiť algoritmy a odstraňovať chyby. Tento výkonový test meria zručnosť modelu pri porozumení a generovaní správneho a efektívneho kódu v rôznych programovacích jazykoch.
- GSM8K – Grade-School Mathematics 8K je výkonový test zameraný na hodnotenie matematických schopností modelu. Zahŕňa rôzne matematické úlohy na úrovni základnej školy, ako sú aritmetika a algebra. Skóre GSM8K udáva, ako dobre dokáže model vykonávať výpočty, chápať matematické pojmy a presne riešiť problémy.
- HellaSwag – referenčný test určený na hodnotenie rozumového uvažovania jazykových modelov prostredníctvom úloh na dokončovanie viet. Poskytuje desať tisíc úloh pokrývajúcich rôzne tematické oblasti.

266 *LLM Leaderboard - Compare Top AI Models for 2024*. [on-line]. [cit. 20. decembra 2024]. Dostupné na internete: <https://yourgpt.ai/tools/llm-comparison-and-leaderboard>

- GPQA – Graduate-Level Google-Proof Q&A Benchmark je test založený na súbore údajov obsahujúcich náročné dvojice otázok s viacerými možnosťami odpovede. GPQA bol vytvorený s pomocou ľudských expertov ako prostriedok na posúdenie výkonnosti LLM v oblasti hraničných znalostí. Tento test je častokrát výzvou aj pre expertov v danej oblasti.
- MMLU – Massive Multitask Language Understanding je referenčný test navrhnutý na meranie znalostí získaných počas predtrénovania vyhodnocovaním modelov výlučne v tzv. *zero-shot and few-shot settings*. Vďaka tomu je tento benchmark náročnejší a podobá sa viac spôsobu, ktorým sú hodnotení ľudia.
- MMMU – Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI je test určený na hodnotenie multimodálnych modelov na masívnych multi-disciplinárnych úlohách vyžadujúcich znalosti predmetov na úrovni vysokej školy a premyslené uvažovanie.[267]
- BBHard Eval – Benchmark for Harder Evaluations in AI je hodnotiaci dataset určený na štúdium pokročilého odvodzovania na základe rozumu, čo je úloha, ktorá je pre ľudí zvyčajne jednoduchá, ale pre stroje náročná.[268]
- MATH – Math Word Problem Solving je test zameraný na riešenie matematických slovných úloh, v ktorých sú významné základné informácie o probléme uvedené skôr v bežnom jazyku ako v matematickom zápise. Keďže väčšina slovných úloh zahŕňa nejaký druh rozprávania, niekedy sa označujú ako príbehové úlohy a môžu sa líšiť množstvom použitého odborného jazyka.[269]

Vo všeobecnosti by testovacie systémy, ktoré sú založené na hodnotení požadovaných vlastností, mali byť škálovateľné (schopné porovnávať rôzne počty modelov navzájom), inkrementálne (schopné vyhodnocovať výkon modelov pomocou relatívne malého počtu pokusov) a vytvárajúce jedinečné poradie (schopné poskytovať jedinečné poradie

267 YUE, X., NI, Y., ZHANG, K. et al. *Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI*. [on-line]. [cit. 20. decembra 2024].
Dostupné na internete: <https://mmmu-benchmark.github.io/>

268 WALKER II, S.M. *BBHard Eval: Pushing the Limits of AI Understanding*. [on-line]. [cit. 20. decembra 2024].
Dostupné na internete: <https://yourgpt.ai/tools/llm-comparison-and-leaderboard>

269 *Math Word Problem Solving*. [on-line]. [cit. 20. decembra 2024].
Dostupné na internete: <https://paperswithcode.com/task/math-word-problem-solving>

pre všetky modely a vzhľadom na ľubovoľné dva modely schopné určiť, ktorý má lepšie výsledky, resp. výsledky sú vyrovnané).[270]

Na základe veľkého množstva rôznych súborov údajov a metrík pokrývajúcich rôzne aspekty generovania môžeme vo všeobecnosti rozdeliť testy do nasledovných kategórií:[271]

- rozumové, logické a programovacie schopnosti:
  HellaSwag, BoolQ, PIQA, SocialIQA, TriviaQA, Natural Questions, ARC-c, ARC-e, WinoGrande, BIG-Bench Hard, DROP, AGIEval, MMLU, MATH, GSM8K, GPQA, MMLU (Pro), MBPP, HumanEval, MMLU (Pro COT), FrontierMath,...
- viacjazyčné schopnosti:
  MGSM, Global-MMLU-Lite, Belebele, WMT24++ (ChrF), FloRes, XL-Sum, XquAD,...
- multimodálne schopnosti:
  COCOcap, DocVQA (val), InfoVQA (val), MMMU (pt), TextVQA (val), RealWorldQA, ReMI, AI2D, ChartQA, ChartQA (augmented), VQAv2, BLINK, OKVQA, TallyQA, SpatialSense VQA, CountBenchQA,...

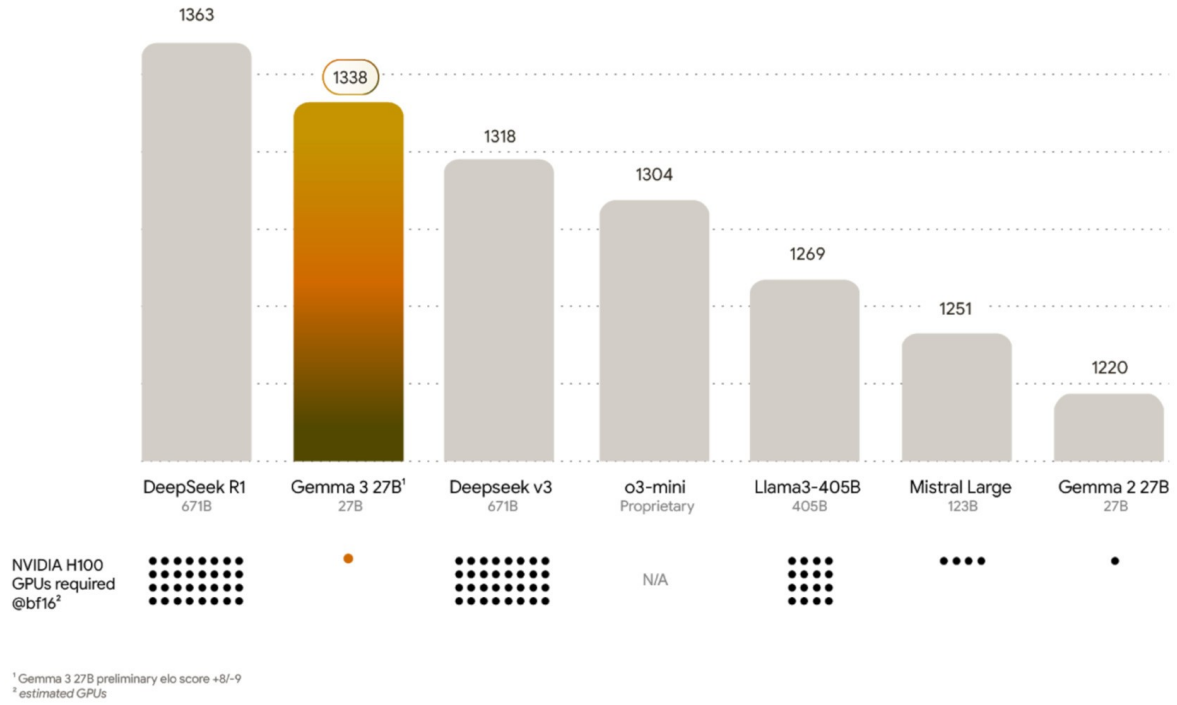
Okrem uvedených testov sa bežne môžeme stretnúť s tzv. systémom hodnotenia Elo score (Elo rating system), ktorý je založený na výpočte relatívnej úrovne zručností hráčov v hrách s nulovým súčtom, ako je šach alebo elektronické športy (video hry), pričom poskytuje rýchly náhľad na približnú výkonnosť generatívneho systému v porovnaní s ostatnými. Príklad porovnania výkonu moderných systémov genAI v Elo score je uvedený na obrázku č. 8.[272]

---

270 ZHENG, L., SHENG, Y., CHIANG, W.L. et al. *Chatbot Arena: Benchmarking LLMs in the Wild with Elo Ratings*. [on-line]. [cit. 13. marca 2025].
Dostupné na internete: <https://lmsys.org/blog/2023-05-03-arena/>

271 Ide o testy bežne používané pre porovnávanie systémov genAI na začiatku roka 2025. *gemma3*. [on-line]. [cit. 12. marca 2025].
Dostupné na internete: <https://ollama.com/library/gemma3:27b>

272 ZHENG, L., SHENG, Y., CHIANG, W.L. et al. *Chatbot Arena: Benchmarking LLMs in the Wild with Elo Ratings*. [on-line]. [cit. 13. marca 2025].
Dostupné na internete: <https://lmsys.org/blog/2023-05-03-arena/>



Obr. č. 8: Porovnanie výkonu systémov genAI v Elo score pri uvedení Gemma3.[273]

Množstvo testov a ich vývoj poukazujú na komplexnosť problematiky generatívnych systémov AI a zložitosť vyhodnocovania ich výkonu, skutočných schopností a potenciálu.

Skladba a obsah jednotlivých testov vyjadrujú testovanie genAI:

- v optike ich smerovania k uvažujúcim systémom a schopnostiam čoraz viac emulovať ľudské myslenie;
- v úzko špecializovaných schopnostiach zameraných na konkrétne oblasti nasadenia;
- v snahe o dosiahnutie čo najlepších výsledkov z pohľadu kvality výstupov, pravdivosti, eliminácie halucinácií a predsudkov, ochrany dát a osobných údajov.

Samostatnú kapitolu tvoria testy na odolnosť (bezpečnosť, robustnosť a spoľahlivosť), interpretovateľnosť a vysvetliteľnosť. Tejto téme sa budeme venovať v 2. kapitole.

273 **Na obrázku je dôležitý nielen dosiahnutý výkon systému AI, ale i podmienky**. Napr. Gemma3 dosahuje vynikajúce výsledky na obyčajnom hardvérovom vybavení (len 1xGPU) a „len“ s 27 miliardami parametrov. Ak si to porovnáme s výkonom a vybavením ostatných modelov, je to špičkový výsledok. *gemma3*. [on-line]. [cit. 12. marca 2025].
Dostupné na internete: <https://ollama.com/library/gemma3:27b>

V rámci úsilia o dosiahnutie silnej a všeobecnej umelej inteligencie (AGI) si osobitnú pozornosť zasluhujú testy, ktoré vychádzajú z referenčného testu „Abstract and Reasoning Corpus for Artificial General Intelligence“ (ARC-AGI) zameraného na meranie efektívnosti získavania zručností umelej inteligencie pri neznámych úlohách.[274]

V súčasnosti viaceré modely implementujúce rozumné uvažovanie dosahujú solídne výsledky v ARC testoch, čo mnohých privádza k presvedčeniu, že uvažujúce modely sa približujú ku skutočnej umelej inteligencii na úrovni človeka (HLAGI).[275]

Toto presvedčenie je však klamlivé, nakoľko AGI-ARC test je zjavne len nevyhnutný, a nie postačujúci test pre HLAGI. Skutočný systém AGI nie je definovaný len tým, že má všeobecné znalosti a bežné schopnosti podobné typickému človeku, ale aj napríklad tým, že má schopnosť zovšeobecňovať na základe skúseností na veľmi odlišné situácie, podobne ako človek.[276] Vyjadrené terminológiou LLM z kapitoly 1.8.1. – generatívne systémy prišli k svojej všeobecnosti znalostí a schopností veľmi zaujímavým druhom učenia, ktoré zahŕňa extrapoláciu pomerne malej vzdialenosti z pomerne veľkého objemu informácií. AGI podobná človeku tiež zahŕňa niečo z tohto učenia, ale zároveň musí disponovať aj inými druhmi učenia, ako napríklad schopnosťou niekedy efektívne kognitívne preskočiť oveľa väčšiu vzdialenosť z malého množstva informácií. **Javí sa, že tento radikálnejší druh „zovšeobecňovania z historickej distribúcie“ je podľa mnohých matematických teórií učenia a kognitívnej vedy spojený s našou schopnosťou vytvárať a používať abstrakcie, a to spôsobom, ktorý súčasné neurónové siete transformerov nedokážu** (z časti je to – aj keď nie celkom presne – ako rozdiel medzi schopnosťou vymyslieť maliarsky štýl Salvadora Dalího a schopnosťou lacno, rýchlo a flexibilne kopírovať maliarsky štýl Salvadora Dalího).[277] Týmto však už atakujeme aj problematiku metakognitívnych schopností, ktoré – ako sme uvádzali v časti *Komplexné rozhodovanie generatívnych systémov* – súčasné generatívne systémy s ich

274 CHOLLET, F. *On the Measure of Intelligence*. [on-line]. [cit. 12. marca 2025]. Dostupné na internete: <https://doi.org/10.48550/arXiv.1911.01547>

275 Prvý z modelov so solídnymi výsledkami v teste ARC-AGI bol model o3 od OpenAI. Výsledky testovania však boli trochu rozporuplné – v niektorých ohľadoch model o3 dosahoval značne nadľudské výsledky, avšak v iných bol veľmi pod schopnosťami človeka.

276 Ďalšie aspekty a schopnosti viažuce sa na skutočnú AGI diskutujeme v 5. kapitole.

277 GOERTZEL, B. *We clearly have not achieved Human-Level AGI*. [on-line]. [cit. 12. marca 2025]. Dostupné na internete: <https://x.com/bengoertzel/status/1878478650548588990>

„ilúziou myslenia“ určite nemajú.

V nadväznosti na ARC-AGI je vhodné spomenúť i **Humanity's Last Exam** (HLE),[278] ktorý podľa výročnej správy Stanford HAI's AI Index 2025 patrí k najnáročnejším súčasným testom.[279] HLE bol vyvinutý v reakcii na existujúce populárne testy umelej inteligencie, ktoré dosiahli „saturáciu“ (ich použitie pre najnovšie systémy AI už nebolo dostatočné). **Pre súčasne pokročilé systémy genAI ide o jeden z najrelevantnejších testov**.

V súčasnosti, t.j. v čase uvádzania sofistikovaných multi-agentových, hybridných, omni-modálnych, ... veľkých modelov genAI, ktoré sa pýšia výbornými výsledkami v uvedených testoch, sa viac a viac potýkame s relevantnosťou výsledkov testovania. Testované modely totižto vykazujú veľkú disproporciu vo výsledkoch svojej činnosti: na jednej strane dosahujú skutočne nadľudské výsledky, no na druhej dokážu byť extrémne podpriemerné. Vo viacerých prípadoch existuje aj reálny predpoklad, že súčasťou ich trénovania a prípravy bola optimalizácia pre dosiahnutie výborných výsledkov v konkrétnych testoch.[280]

Záverom...

Viaceré aspekty a nádejné smery vylepšovania, ktoré sme v tejto kapitole uviedli, vyjadrujú veľký potenciál generatívnych systémov AI, súvisiaci nielen s rastúcou schopnosťou

278 Ide o bechmark „na pomyselnej hranici ľudského poznania“ so širokým tematickým záberom. Je to štandardizovaný pokročilý test jazykových modelov vyhodnocujúci ich výkon v rôznych špecifických oblastiach (porozumenie jazyka, generovanie výstupov, uvažovanie,...).
HLE pozostáva z 2 500 otázok vo verejne dostupnej sade, pričom klasifikuje otázky do nasledujúcich oblastí: matematika (41%), fyzika (9%), biológia/medicína (11%), humanitné/spoločenské vedy (9%), informatika/umelá inteligencia (10%), strojárstvo (4%), chémia (7%) a ostatné (9%).
Približne 14% otázok vyžaduje schopnosť porozumieť textu aj obrázkom, t. j. multimodalitu. 24% otázok je vo forme výberu z viacerých možností, zvyšok tvoria otázky s krátkou odpoveďou a otázky vyžadujúce presné odpovede. Udržuje sa aj sada na testovanie nadmerného prispôsobovania sa (overfitting).
PHAN, L., GATTI, A., HAN, Z. et al. *Humanity's Last Exam*. [on-line]. Center for AI Safety, Scale AI, 2025. [cit. 12. júla 2025].
Dostupné na internete: <https://doi.org/10.48550/arXiv.2501.14249>

279 MASLEJ, N. et al. *The AI Index 2025 Annual Report*. [on-line]. Institute for Human-Centered AI, 2025. [cit. 12. júla 2025].
Dostupné na internete: <https://hai-production.s3.amazonaws.com/files/hai_ai_index_report_2025.pdf>

280 V konečnom dôsledku platí: **skutočným testom schopností systému AI bude jeho fungovanie v reálnom nasadení, resp. v reálnom svete**.

emulovať myslenie a vykonávať zložité zadania, ale i s kvantom dát, na ktorých sa učia.

Parametre veľkých modelov, ktoré dosahujú bilióny natrénovaných hodnôt, by sme v prenesenom zmysle slova mohli považovať za stratovú kompresiu sveta – takého, aký je prezentovaný na internete, sociálnych sieťach, vo vedeckých prácach, zdigitalizovanom umení a pod. Keďže matematicky ide o veľkú podobnosť medzi predikciou a kompresiou, súbor natrénovaných parametrov, ktoré majú zakomponovaných veľa informácií o reálnom svete a ich fungovaní tak, ako je to zachytené v trénovacích dátach, by sme nadnesene mohli nazvať **„zazipovaným“ obrazom sveta** s obrovským potenciálom pre toho, kto ho vlastní a využíva.[281]

Preto tie najlepšie, najväčšie a najsofistikovanejšie generatívne modely s (v tejto kapitole) predstaveným potenciálom rozvoja môžeme považovať za strategické technológie, ktoré majú potenciál meniť svet – k lepšiemu, ale i k horšiemu...

Je však možné, že stojíme na konci jednej éry rozvoja generatívnych systémov AI. Ak zoberieme do úvahy analógiu s trénovaním modelov AlphaGo až Alpha Zero, t.j. s hľadaním spôsobu ako ďalej trénovať, uvedomujeme si, že **nám dochádzajú kvalitné tréningové údaje**, na ktorých by sme mohli systémy AI skutočne trénovať, keďže sme už takmer všetko dostupné a použiteľné zo súčasného virtuálneho sveta využili.[282]

Podľa Ilyu Sutskevera, ktorý uvedenú skutočnosť zhrnul do vyjadrenia „**dáta sú fosílnym palivom AI** – dosiahli sme maximum dát“, existuje viacero možností ďalšieho vývoja nových systémov umelej inteligencie:[283]

- rozmach využívania agentov AI, čo síce neznamená príchod úplne nových technológií, ale posun smerom k samostatným systémom AI (ako sme diskutovali

281 KARPATHY, A. *Intro to Large Language Models*. [on-line]. [cit. 13. februára 2024]. Dostupné na internete: <https://youtu.be/zjkBMFhNj_g>
Skúsenosť „zazipovaného“ informačného obrazu sveta si možno osobitne uvedomiť pri použití kvalitných lokálnych modelov genAI. Ich kreatívne využívanie mnohokrát pripomína prácu s takýmto informačným obrazom sveta vo „vreckovom“ vydaní:-)

282 SUTSKEVER, I. *Sequence to sequence learning with neural networks: what a decade*. [on-line]. [cit. 17. decembra 2024]. Dostupné na internete: <https://www.youtube.com/watch?v=1yvBqasHLZs>

283 SUTSKEVER, I. *Sequence to sequence learning with neural networks: what a decade*. [on-line]. [cit. 17. decembra 2024]. Dostupné na internete: <https://www.youtube.com/watch?v=1yvBqasHLZs>

osobitne v častiach o agentoch AI a hybridných systémoch);

- využívanie syntetických údajov, t.j. hľadanie nových spôsobov, ako generovať tréningové dáta pre vylepšovanie modelov AI;
- ďalší pokrok smerom ku komplexnému rozhodovaniu genAI (myslenie System2), či už vylepšovaním pokročilého spracovania promptu v strome možností alebo prechodom k logickému mysleniu (na spôsob napr. logiky prvého rádu).

K úvahe Ilyu Sutskevera by sme mohli doplniť ešte jeden súčasný trend, resp. spôsob, ako sa dostať k ďalším reálnym dátam – ide o **zdroje, ktoré sú prepojené s dátami konkrétnych používateľov a ktoré bežne nie sú dostupné**. Napr. integrácia služieb AI do operačných systémov (MS Windows, Android,...), komunikátorov a kolaboratívnych nástrojov (WhatsApp, iMessages, Teams, Zoom,...), aplikácie pre narábanie s médiami (fotografie a video záznamy v mobiloch) a pod. Ide o dáta, ktoré doteraz boli súčasťou súkromných, niekedy aj E2E šifrovaním chránených komunikácií a lokálnych, prípadne súkromných cloudových úložísk.[284]

Viaceré, v tejto kapitole načrtnuté možnosti vývoja systémov AI majú evolučný potenciál smerovania ku skutočnej umelej inteligencii (AGI). Súčasne sme však uviedli výhrady, prečo a v čom sú tieto možnosti nedostatočné a bez ďalšieho zásadného pokroku nás ku skutočnej inteligencii neprivedú.

Čo sa týka blízkej budúcnosti, Sutskever je presvedčený, že budúce systémy umelej inteligencie sa budú kvalitatívne líšiť od dnešných modelov v niekoľkých kľúčových ohľadoch:[285]

---

284 S konkrétnymi pokusmi sa v poslednom čase „roztrhlo vrece“, napr.:
Gemini can read and act on your phone, messages, and WhatsApp — even if you've turned "Gemini Apps Activity" off. [on-line]. [cit. 28. júna 2025].
Dostupné na internete: <https://x.com/WireMin/status/1937766318947844220>
Osobitne pre tzv. personalizované systémy AI a osobných agentov je treba umožniť **prístup k osobným, častokrát dôverným a citlivým dátam, t.j. prakticky ku kompletnej dátovej stope osoby**.
Powering Safe, Personalized Galaxy AI Experiences. [on-line]. [cit. 10. júla 2025].
Dostupné na internete: <https://news.samsung.com/global/your-privacy-secured-inside-the-tech-powering-safe-personalized-galaxy-ai-experiences>

285 SUTSKEVER, I. *Sequence to sequence learning with neural networks: what a decade*. [on-line]. [cit. 17. decembra 2024].
Dostupné na internete: <https://www.youtube.com/watch?v=1yvBqasHLZs>

- budú mať skutočné schopnosti: nebudú sa len riadiť pokynmi, ale budú sledovať svoje vlastné ciele;
- budú skutočne uvažovať: pôjde o prechod od porovnávania vzorov k logickému mysleniu;
- dokážu sa lepšie učiť: budú schopné chápať pojmy na základe obmedzených údajov;
- budú mať sebauvedomenie: budú mať vnútorný model seba samého;[286]
- budú nepredvídateľné: čím viac budú uvažovať, tým ťažšie sa budú dať predvídať.

Viaceré z uvedených kľúčových vlastností atakujú tzv. koncepčné prelomy na ceste k AGI, ktoré diskutujeme v 5. kapitole a v ktorej si aj ukážeme, že problematika AGI je oveľa komplexnejšia.

Symbolickým zavŕšením našej úvahy o generatívnych technológiách umelej inteligencie by mohol byť – obrazne povedané – návrat k človeku, nakoľko aktuálne výskumy poukazujú na **ľudskú adaptáciu, ktorá odomyká potenciál systémov genAI**. Viaceré oblasti implementácie generatívnych systémov sprevádzala falošná premisa: ako budú modely umelej inteligencie stále výkonnejšie, vymizne potreba ľudských zručností pri ich riadení.[287]

Avšak napríklad štúdia Eamana Jahaniho a kol. s názvom „As Generative Models Improve, People Adapt Their Prompts“ (Ako sa generatívne modely zlepšujú, ľudia prispôsobujú svoje vstupy) prináša presvedčivé experimentálne dôkazy, ktoré to vyvracajú.[288] Prostredníctvom rozsiahleho experimentu výskumníci dokázali, že **vzťah medzi ľuďmi a pokročilou umelou inteligenciou nie je založený na znižovaní ľudského vkladu, ale skôr na „koadaptívnom tanci“, v ktorom ľudia intuitívne rozvíjajú svoju komunikáciu a zručnosti, aby odomkli nové schopnosti vyspelých**

---

286 **Sebauvedomenie vyjadrené ako schopnosť mať vnútorný model seba samého je síce elegantné, resp. popisne správne, ale súčasne extrémne zjednodušujúce a bez reálnej cesty ponajprv k pochopeniu a následne ku komplexnému technologickému uchopeniu tohoto fenoménu.**

287 Niektorí technológovia predpovedajú budúcnosť, v ktorej bude umelá inteligencia taká zručná v interpretácii zámerov, že špecializovaná prax „prompt engineering“ (riadenie umelej inteligencie) sa stane zastaranou.

288 JAHANI, E., MANNING, B.S., ZHANG, J. *As Generative Models Improve, People Adapt Their Prompts*. [on-line]. [cit. 12. októbra 2025].
Dostupné na internete: <https://arxiv.org/abs/2407.14333v2>

modelov. Hlavné zistenia štúdie – že používatelia píšu zložitejšie prompty pre lepšie modely a že táto adaptácia správania predstavuje takmer polovicu nárastu výkonu – naznačujú, že prompty zostanú kľúčovým mechanizmom pre využitie plného potenciálu súčasných systémov genAI. Dokonca sa ukázalo, že ak ľudský vklad bol nahradený pokročilými asistenčnými službami (tiež realizovanými pomocou AI;-), schopnosť efektívne využívať schopnosti vyspelých modelov dramaticky poklesla.[289]

I táto štúdia veľmi presvedčivo vyvracia tvrdenie, že ľudské zručnosti budú v ére pokročilej umelej inteligencie irelevantné, a tiež načrtáva obraz „dynamiky synchronizovaného kroku“, v ktorej ľudská vynaliezavosť a technologický pokrok postupujú spoločne.[290] **Ako sa modely zlepšujú, vytvárajú nový potenciál a ľudia prispôsobujú svoj jazyk a stratégie, aby tento potenciál využili**. Pokrok generatívnej umelej inteligencie zďaleka neznehodnocuje prompting, ale javí sa, že ho ešte viac zdôrazňuje, pretože je primárnym rozhraním, prostredníctvom ktorého sa objavujú a využívajú nové schopnosti. Štúdia presvedčivo argumentuje, že prompting zostane kľúčom, ktorý odomkne ďalšiu generáciu systémov genAI, a **zručnosť v jeho používaní bude rozhodujúcim faktorom, určujúcim schopnosť profitovať z transformačnej sily týchto nástrojov**.[291]

289 PEREZ, C.E. *The Lock-Step Dance: How Human Adaptation Unlocks AI's Potential*. [on-line]. [cit. 12. októbra 2025].
Dostupné na internete: <https://x.com/IntuitMachine/status/1959251441924608074>

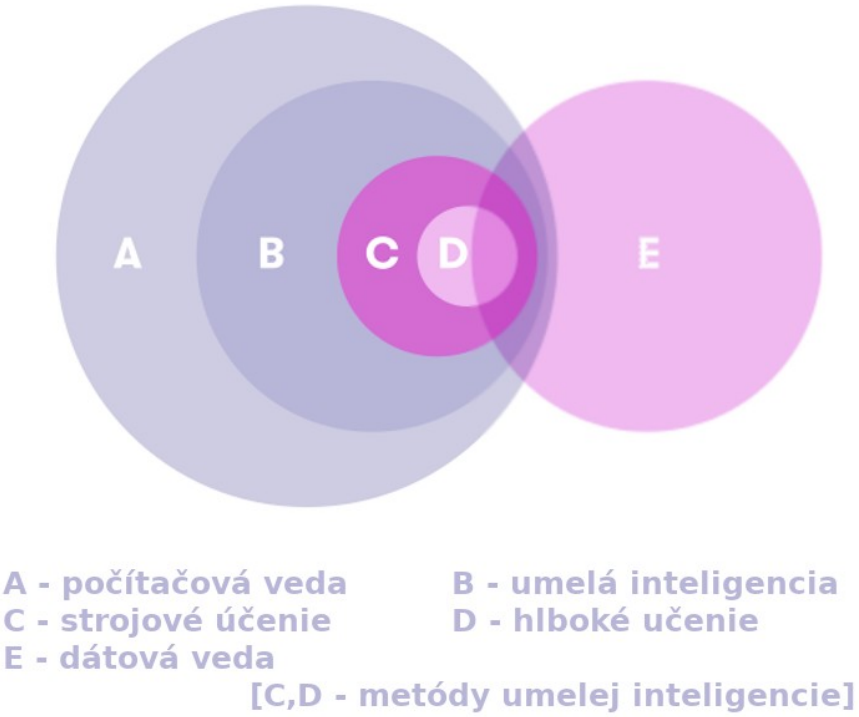
290 Viaceré testy (konané napríklad aj na UK v Bratislave) poukazujú na **závislosť kvality výstupov od formy, spôsobu a kvality zadávaných promptov**. Ten istý model genAI dáva výstupy rôznej kvality v závislosti od formulácie vstupného promptu, resp. celého zadania.
Napr. kvalita výstupu môže rásť s nasledovnými zmenami v prompte: jednoduché zadanie úlohy -> zadanie so žiadosťou, aby model opísal kroky, ktoré pri riešení vykonáva -> zadanie žiadajúce, aby model riešil úlohu komunikáciou viacerých expertov,...
Z výsledkov dedukujeme, že **rôzne formy zadania aktivujú iné časti dátového priestoru modelu, v ktorých akoby platili iné geometrické vzťahy (v rámci vektorizácie a transformačných matíc), určujúce kvalitu a komplexnosť odpovedí**. Je to v zásade vyjadrenie skutočnosti, že miesto cieleného myslenia sa model na základe konkrétnych promptov len pohybuje v rôznych častiach svojho dátového priestoru.
TAKÁČ, M. *Moderná umelá inteligencia: Čo je v čiernej skrinke?* [on-line]. Prednáška na FMFI UK, 3.10.2025. [cit. 12. októbra 2025].
Dostupné na internete: <https://fmph.uniba.sk/detail-novinky/back_to_page/fakulta-matematiky-fyziky-a-informatiky-uk/article/seminar-odi-martin-takac-3102025/>

291 PEREZ, C.E. *The Lock-Step Dance: How Human Adaptation Unlocks AI's Potential*. [on-line]. [cit. 12.

1.9. Striedanie ročných období a najbližšia predpoveď počasia

Skúsme spraviť malú retrospektívu vývoja umelej inteligencie. Roky po Dartmouthskom seminári ubiehali a veľmi optimistické predpovede o pokroku vo vývoji systémov umelej inteligencie sa nenapĺňali.[292] S vyčerpaním potenciálu konkrétnych vtedajších poznatkov v oblasti AI a schopností dobovej techniky preto veľmi rýchlo prichádzalo k strate nadšenia i utlmeniu investícií do tejto oblasti.

Vývoj bol udržiavaný v zásade len v rámci základného výskumu viacerých univerzitných centier a pokroku v príbuzných oblastiach (napr. robotika a kybernetika, výpočtová technika, dátová a počítačová veda, v rámci priemyslu a vojenského vývoja), z čoho ťažila aj veľmi vágne ohraničená oblasť umelej inteligencie, ako sme uviedli na konci kapitoly 1.2. Vzťah medzi dátovou vedou, počítačovou vedou a umelou inteligenciou ako vedným odborom vyjadruje schematický obrázok č. 13.



Obr. č. 13: Vzťah medzi dátovou vedou, počítačovou vedou a umelou inteligenciou.

I napriek útlmu ovocím základného výskumu a pokroku v príbuzných oblastiach bol posun v hľadaní nových metód a prístupov k riešeniu problémov umelej inteligencie. Každý väčší

---

októbra 2025].

Dostupné na internete: <https://x.com/IntuitMachine/status/1959251441924608074>

292 Viac o optimistických predpovediach a predikciách hlavných protagonistov uvádzame v poznámke č. 74.

posun sa následne stával štartérom nového entuziazmu a záujmu o umelú inteligenciu, prichádzali nové pozitívne predikcie, ktorých dôsledkom bol i ďalší prísun investícií a podpory pre túto oblasť. A po vyčerpaní potenciálu daného posunu znovu nastával útlm...

Toto – až takmer do poslednej dekády viac menej pravidelné – **striedanie aktívneho rozvoja a útlmu v oblasti umelej inteligencie sa nazýva aj striedanie ročných období**:

- jar umelej inteligencie (AI spring) – obdobie prekotného rozvoja, ktoré štartuje novými ideami a pokrokom vo výskume, čoho ovocím sú predikcie prevratného pokroku vo vývoji systémov AI, častokrát sprevádzané mediálnym boomom a následným prílevom štátnych dotácií i rizikového kapitálu do akademického bádania i komerčných startupov.
- zima umelej inteligencie (AI winter) – prevratný pokrok neprichádza, výsledky síce sú, no vôbec nespĺňajú extrémne očakávania. Prílev financií a kapitálu ustáva, startupy krachujú a akademický výskum sa spomaľuje a obmedzuje.

Uvedený vzor sa vo vývoji umelej inteligencie v rozličných podobách opakoval v päť až desaťročných cykloch.[293]

S veľkým pokrokom vo vývoji algoritmov vychádzajúcich zo štatistických a pravdepodobnostných teórií, ktoré umožnili rozvoj strojového učenia, v súčasnosti postupne prišlo – až na niekoľko špecifických oblastí – k odmietnutiu symbolických systémov AI[294] a veľkému nástupu neurónových sietí a strojového učenia. V posledných dvoch desaťročiach tak striedanie ročných období reálne prebieha už len v rámci vlastných cyklov vývoja subsymbolických systémov AI, keďže viackrát sa ukázalo, že na riešenie skutočných problémov reálneho sveta sú aj moderné systémy strojového učenia prikrátke...

V poslednej dekáde možno diskutovať o ďalšom vývoji umelej inteligencie i z pohľadu týchto vývojových cyklov.

---

293 Por. MITCHELL, *Artificial Intelligence*, s. 34.

Autorka dokonca spomína, že v čase ukončenia jej vysokoškolského štúdia (1990) sa vývoj AI nachádzal tak hlboko „v zimnom období“, až je radili, aby do profesného životopisu v rámci žiadostí o zamestnanie umelú inteligenciu vôbec neuvádzala.

294 Odmietnutie bolo často sprevádzané dešpektom zhmotneným do označenia symbolickej AI ako GOFAI (podľa Douglasa Hofstadtera ide o skratku pre good old old-fashioned AI).

Umelá inteligencia v súčasnosti zažíva nebývalý rozmach, dynamický vývoj, extrémny rozsah praktického aplikovania jej systémov, špičkové zabezpečenie výskumu a vývoja, spoločenskú akceptáciu i prehnané mediálne pokrytie a očakávania. Vyzerá to skoro ako definícia jarnej fázy AI.

Na druhej strane však treba povedať, že aktuálne veľké pokroky sú len v malej miere dôsledkom – nazvime to – revolučných zmien v AI (môžme takto čiastočne uvažovať o generatívnych systémoch AI a ich celospoločenskej akceptácii), keďže vo veľkej miere ide len o evolučný vývoj existujúcich metód a ich sofistikovaných kombinácií, technologickú podporu z iných oblastí (hlavne rozvoj informačných a komunikačných technológií, osobitne vysokovýkonných systémov), dostupnosť extrémne veľkých datasetov a vytrénovaných modelov genAI, atď. Prevratný pokrok v podstatných aspektoch, ktorými myslíme vytvorenie všeobecnej umelej inteligencie „aspoň“ na úrovni HLAGI, však neprichádza. Ak by sa teda rozvoj súčasných systémov AI (predovšetkým genAI) spomalil alebo vyčerpal, mohlo by to skôr pripomínať prvé zimné mrazy...

Keďže miera akceptácie a adaptácie systémov AI v modernej spoločnosti presiahla hranicu, za ktorou je ťažké si predstaviť návrat späť (čím by mohlo byť napríklad zlyhanie generatívnych systémov alebo zmena priorít rozvoja vedomostnej spoločnosti a Industry 4.0) a keďže v niektorých aspektoch trochu plochá krivka základného výskumu a vývoja systémov AI je súčasťou permanentného technologického pokroku, ktorý celkovo v oblasti rozvoja vedy, techniky, objemu dát a znalostí dosahuje takmer exponenciálny rast,[295] dovolíme si predpokladať, že **na aktuálny a budúci vývoj v oblasti umelej inteligencie už nebude možné v plnej miere aplikovať striedanie jarných a zimných období. Výskum a vývoj systémov AI bude neustále pod tlakom prakticky všetkých spoločenských oblastí, v ktorých sa AI v súčasnosti využíva**.

Pozorný čitateľ určite vníma, že spochybňujúc pokračovanie vývojových cyklov sme skĺzli len do oblasti ANI, t.j. k úzko špecializovaným systémom (slabej) umelej inteligencie, ktorým sa v súčasnosti neuveriteľne darí.

Inak to však môže byť s AGI, teda so skutočnou umelou inteligenciou, ktorá sa stáva

295 ŠANTAVÝ, *Analýza virtuálneho sveta v kontexte náboženskej formácie, pôsobenia a života Cirkvi,* [online], s. 20, 25, 31, 34. [cit. 7. januára 2022]. Dostupné na internete: <https://peter.santavy.cloud/data/uploads/docs/analyza_virtualneho_sveta_v_kontexte_nabozenskej_formacie_posobenia_a_zivota_cirkvi.pdf>

definitívnou métou súčasného základného výskumu v oblasti AI.[296] Tu si vývoj nedovolíme predpovedať, keďže na jednej strane principiálny prelom vo vývoji metód umelej inteligencie nenastáva, na strane druhej však vylepšovanie a kombinácia súčasných techník, medzi ktoré patrí aj nástup generatívnych systémov AI – osobitne modely s rozumným a kolaboratívnym uvažovaním, systémy AI agentov, extrémne veľké datasety, výpočtové systémy s výkonom prispôsobeným priamo pre aplikácie umelej inteligencie i všeobecný vedomostný a výskumný potenciál môžu v kombinácii s novými objavmi veľmi ľahko viesť aj k prípadnej skokovej zmene a vývojovému prelomu, resp. k jednému alebo viacerým koncepčným prelomom, ktoré diskutujeme v 5. kapitole.

V rámci takmer sedemdesiatich rokov bádania a snáh o vytvorenie všeobecnej umelej inteligencie mnohí vedci a výskumníci viackrát takmer opustili ideál AGI. V súčasnosti – keď sa debata o umelej inteligencii stala súčasťou hlavného spoločenského prúdu – sa mnohým tento ideál javí ako samozrejmý. Predpokladáme, že len niekoľko rokov bude stačiť, aby sme videli, ktoré tendencie prevažujú: či optimizmus pretrváva, alebo fenomén AGI sa v rámci pokračujúceho výskumu bude javiť ako oveľa sofistikovanejší, komplexnejší a náročnejší nielen na realizáciu, ale hlavne na uchopenie a pochopenie.

1.10. Jednoduché veci sú ťažké

Nadväzujúc na predchádzajúcu kapitolu treba povedať, že v určitej miere nepôjde o nič nové, keďže súčasťou každej zimy umelej inteligencie bolo zistenie, akým veľkým problémom pre systémy AI je schopnosť realizácie niektorých vecí. Päťdesiat rokov od konferencie v Dartmouthe to John McCarthy vyjadril slovami: „AI bola ťažšia, ako sme si mysleli“.[297] A Marvin Minsky už oveľa skôr poznamenáva, že výskum v oblasti umelej inteligencie odkryl **dôležitý paradox: „Jednoduché veci sú ťažké“[298], keďže riešenie**

296 Filozofi umelej inteligencie, Vincent Müller a Nick Bostrom zverejnili v roku 2013 prieskum vykonaný medzi výskumníkmi AI, v ktorom mnohí vyjadrili 50% šancu na dosiahnutie umelej inteligencie na úrovni človeka do roku 2040.

MÜLLER, V. C., BOSTROM, N. *Future Progress in Artificial Intelligence: A Survey of Expert Opinion*. In: *Fundamental Issues of Artificial Intelligence. Cham.* Switzerland: Springer International, 2016, 555-72.

Treba však poznamenať, že s nástupom a hlavne aktuálnym rozvojom genAI sa predikcie skracujú – mnohí s nadšením očakávajú príchod HLAGI už do roku 2030.

297 MOEWES, C., NÜRNBERGER, A. *Computational Intelligence in Intelligent Data Analysis*. New York: Springer, 2013, s. 135.

298 MINSKY, M. *Society of Mind*. New York: Simon & Schuster, 1986, s. 29.

úloh, ktoré sa človeku zdajú byť jednoduché, je pre umelú inteligenciu veľmi náročné. A naopak to, čo je pre človeka extrémne náročné, dokážu systémy AI zvládať veľmi dobre.[299]

Pôvodné ciele vývoja umelej inteligencie – počítače, ktoré dokážu s nami komunikovať ľudskou rečou, vedia popísať, čo „vidia“ kamerou, „chápu“ koncepty na základe len niekoľkých príkladov – to všetko sú veci, ktoré malé dieťa poľahky zvláda, no pre AI sú oveľa ťažšie, ako napr. komplexná diagnostika choroby, víťazstvo v šachu alebo v Go nad najlepšími ľudskými šampiónmi alebo riešenie sofistikovaného algebraického problému.[300] [301]

Minský to vyjadril slovami: „Vo všeobecnosti si najmenej uvedomujeme to, čo naša myseľ robí najlepšie.“[302] A Sparsh Chadha dodáva: „Stroje ... dokážu vykonávať len úlohy, na ktoré boli vytrénované. Vďaka našej schopnosti dynamického myslenia a zdravému rozumu výrazne prevyšujeme stroje. Bez ohľadu na to, ako veľmi natrénujeme naše modely, na koľkých tréningových cykloch trénujeme naše stroje, stále existuje priestor neistoty, ktorý by sa dal človekom veľmi jednoducho vyriešiť, ale umelá inteligencia by zlyhala, pretože jej chýba **schopnosť zdravého rozumu**.“[303]

Ide o skúsenosť sprevádzajúcu celé desaťročia vývoja umelej inteligencie, ktorá nám pripomína, ako komplexná a subtílna je ľudská myseľ a čo všetko je na ceste k AGI ešte pred nami.

---

299 Kognitívny vedec Steven Pinker to vyjadruje celé: **„The hard things are easy, but the easy things are hard.“**
TROTT, D. *The hard things are easy, but the easy things are hard*. [on-line]. [cit. 8. januára 2022]. Dostupné na internete: <https://www.campaignlive.com/article/hard-things-easy-easy-things-hard/1498154>

300 MITCHELL, *Artificial Intelligence*, s. 33-34.

301 Ide o tzv. **Moravcov paradox**, ktorý bol jedným z výsledkov výskumu Hansa Moravca, Rodney Brooksa a Marvina Minského v osemdesiatych rokoch.
MORAVEC, H. *Mind Children: The Future of Robot and Human Intelligence*. Cambridge, Mass.: Hardvard University Press, 1988.

302 MINSKY, *Society of Mind*, s. 29.

303 CHADHA, S. *“Common Sense” is the Dark Matter of Artificial Intelligence*. [on-line]. [cit. 4. februára 2022].
Dostupné na internete: <https://hackernoon.com/the-dark-matter-of-ai-common-sense-is-not-so-common>

Jednou z podmienok zvládnutia pre človeka jednoduchých, ale pre AI náročných vecí je prechod od systémov, ktoré sa javia ako inteligentné (napr. systém AI na jazykový preklad, systém na detekciu a popis objektov zosnímanej scény a pod.) **k systémom, ktoré sú inteligentné** (AI systém, ktorý rozumie prekladanému textu, systém, ktorý chápe obsah zosnímanej scény a pod.).[304] Rozumieť a chápať súvislosti – to sa už prakticky približujeme k schopnostiam všeobecnej a silnej umelej inteligencie (AGI).

Prakticky vo všetkých strategických oblastiach rozvoja umelej inteligencie je AGI túženým cieľom, keďže jej schopnosti by umožnili riešiť problémy na úplne inej úrovni. V kombinácii s technologickými možnosťami súčasnej informatizácie a automatizácie sa potencionálne schopnosti AGI javia ako úžasné.

Napr. analýza cestnej premávky na základe všetkých dostupných senzorov autonómneho vozidla – teda schopnosť ľudského komplexného vnímania, ktorá by bola spojená s vizuálnymi, zvukovými, radarovými, lidarovými, atď. senzormi a zároveň by na základe chápania aktuálnej dopravnej situácie umožňovala extrémne rýchle a presné reakcie v priamej interakcii s elektronickým ovládaním vozidla.[305]

Tieto predpokladané úžasné možnosti idú tiež ruka v ruke s výzvami, ktoré boli v oblasti rozvoja systémov umelej inteligencie pomerne dlho na okraji záujmu – **etický rozmer činnosti strojov: ako ho definovať, ako ho realizovať**, ako fixovať požadované vlastnosti a ako ich aplikovať do predmetu činnosti daného systému umelej inteligencie**.**

Etika systémov umelej inteligencie je v súčasnosti už pomerne zavedená oblasť výskumu AI. **U AGI však nejde len o etiku vývoja, nasadenia, prevádzkovania a využívania systémov umelej inteligencie, ale priamo o etický rozmer nezávislého a autonómneho dosahovania cieľov v rámci činnosti týchto systémov**.

1.11. Transhumanizmus a umelá inteligencia – tandem i súperi

Koketovanie s myšlienkou AGI a túžba po vytvorení skutočnej, dokonca až uvedomelej umelej inteligencie je často konfrontovaná s inteligenciou človeka. V optike problémov,

304 Pozri poznámku č. 64 a súvisiaci text v kapitole 1.1.

305 Súčasné autonómne systémy sú mnohokrát veľmi pomalé a nedokonalé, ak sa majú rozhodovať napr. v situácii zložitej križovatky (napr. autopilot na bežnej cestnej komunikácii), alebo dostatočne výkonné, avšak bytostne závislé na externých podporných technológiách (napr. špeciálny autopilot taxíkov v čínskom Šen-čene/Shenzhene závislý od miestnej infraštruktúry).

ktoré zahŕňa vývoj AGI a v obavách z disproporcie medzi skutočnou umelou a ľudskou inteligenciou, sa mnoho vedcov a futurológov zaoberá myšlienkami transhumanizmu, t.j. futurologickým a filozofickým konceptom, ktorý pojednáva o možnostiach transformácie človeka prostredníctvom využitia moderných technológií. Je to pohľad do budúcnosti, ktorý je založený na premise, že ľudia vo svojej súčasnej podobe nereprezentujú koniec vývoja, ale iba jeho ranú fázu.[306]

V zásade môžeme uvažovať o dvoch základných rozmeroch vzťahu medzi umelou inteligenciou a transhumanizmom:

- integrácia systémov AI a človeka
- konfrontácia s AI, existenciálne riziko vyplývajúce z pokročilej umelej inteligencie

Zavádzanie prvkov umelej inteligencie zahŕňa aj asistenčné technológie, ktoré zvyšujú kvalitu života, od tých jednoduchých, napr. implementácia prvkov rozšírenej reality (augmented reality) v rámci rôznych nositeľných zariadení (wearables), až po oveľa sofistikovanejšie, napr. pripojenie robotických končatín na nervovú sústavu hendikepovaného človeka[307]. V tomto kontexte sa pre časť výskumu javí atraktívnou

306 Transhumanizmus je futurologický a filozofický koncept, ktorý sa zaoberá možnosťami transformácie človeka prostredníctvom využitia moderných technológií. Ide o filozofické a intelektuálne hnutie, ktoré obhajuje zlepšenie ľudského stavu vývojom a sprístupnením sofistikovaných technológií, ktoré dokážu výrazne zvýšiť dĺžku a kvalitu života i kognitívne schopnosti. Transhumanistickí myslitelia skúmajú potenciálne výhody a nebezpečenstvá vznikajúcich technológií, ktoré by mohli prekonať základné ľudské obmedzenia, ako aj etiku používania takýchto technológií.
MERCER, C., TROTHEN, T. J. *Religion and Transhumanism: The Unknown Future of Human Enhancement*. Praeger, 2014.
Niektorí transhumanisti sa domnievajú, že ľudia sa nakoniec budú môcť premeniť na bytosti so schopnosťami tak výrazne rozšírenými oproti súčasnému stavu, že si zaslúžia označenie post-humánne bytosti (pojednáva o tom samostatný komplexný futuristický smer označovaný ako posthumanizmus).

307 Elon Musk v rámci svojich vizionárskych aktivít (Neuraling Corporation) vytvára **rozhranie medzi mozgom a počítačom pre rozšírenie možností človeka**, ako napr. integrácia robotických končatín, alebo prepojenie ľudského mozgu a umelej inteligencie pre dosiahnutie symbiózy, ktorú by sme poľahky mohli zaradiť do oblasti transhumanizmu.
Neuralink - Breakthrough technology for the brain. [on-line]. [cit. 31. júla 2025].
Dostupné na internete: <https://neuralink.com/technology/>
Neuralink [on-line]. [cit. 5. augusta 2020].
Dostupné na internete: <https://en.wikipedia.org/wiki/Neuralink>

myšlienka prepojiť vhodné systémy z umelej inteligencie s telom a nervovou sústavou človeka.[308] Osobitne, ak na jedne strane dosahujeme pomerne veľké úspechy vo vývoji ANI, no súčasne sa len málo posúvame k dosiahnutiu méty AGI.

Keďže jednou z tém a oblastí transhumanistického výskumu je snaha ochrániť ľudstvo pred existenčnými rizikami,[309] ako je napríklad jadrová vojna alebo zrážka s asteroidom, smerovanie k všeobecnej a uvedomelej umelej inteligencii spojené s rizikami, ktoré jej potenciál môže v sebe ukrývať, viedli transhumanistickú scénu aj k vnímaniu existenciálnych rizík vyplývajúcich z pokročilej umelej inteligencie.[310]

V rámci (nielen) konfrontácie ľudského bytia s AI transhumanistické vylepšovanie zahŕňa rozličné vedné oblasti – genetiku, biológiu a medicínu, psychológiu, ... – ktoré nemusia priamo súvisieť umelou inteligenciou.[311]

Ak by sme len v úzkom kontexte tejto publikácie chceli uchopiť etické výzvy spojené s transhumanizmom[312], išlo by o:

---

308 Treba povedať, že okrem invazívneho prístupu priameho prepojenia, ako je to v prípade Muskovho Neuralingu, rozvíja sa i neinvazívny prístup, ktorý s technologickým rozvojom senzorov a systémov AI v súčasnosti prináša zaujímavé výsledky.

309 „Elon Musk už dlho hovorí, že umelá inteligencia bude musieť ľudské schopnosti skôr rozširovať, než s nimi súťažiť, aby sa vyhla hrozivej budúcnosti.“
FERRIS, R. *Elon Musk thinks we will have to use AI this way to avoid a catastrophic future*. [on-line]. [cit. 24. januára 2022].
Dostupné na internete: <https://www.cnbc.com/2017/01/31/elon-musk-thinks-we-will-have-to-use-ai-this-way-to-avoid-a-catastrophic-future.html>

310 BOSTROM, N. *A history of transhumanist thought*. In: *Journal of Evolution and Technology.* [on-line]. 2005, zv. 14, vyd. 1. [cit. 8. januára 2022].
Dostupné na internete: <http://www.nickbostrom.com/papers/history.pdf>

311 Ide napríklad o vojenské vylepšenia typu „skin-in“, teda o technológie, ktoré modifikujú ľudskú biológiu alebo ju využívajú neštandardným spôsobom, aby človeku zabezpečili schopnosti, ktoré nemá, alebo dostali ich na úroveň ďaleko presahujúcu mieru bežného človeka. Medzi tieto technológie patria kognitívne zosilňovače, metabolická podpora (premena kryštalickej celulózy na glukózu, blokovanie myostatínu, cvičebné pilulky, Peek Soldier Performance, resp. neurochemické látky, hormonálne kokteily), génové terapie, vakcíny proti bolesti a pod.
THURZO, V., Vivoda, M., REIMER, T, *Man and Machine*, Dublin: International scientific board of catholic researchers and teachers in Ireland, 2023. s. 60-130.

312 Hovoríme o výzvach spojených priamo s AI, nie s celým rozsahom problémov týkajúcich sa transhumanizmu.

- problematiku externých technológií AI (tzv. „skin-out“), ktoré zahŕňajú pokročilé nástroje AI, nositeľné zariadenia (waerables) a ich kombináciu vo forme zariadení rozšírenej reality (augmented reality);[313]
- problematiku implantátov, ktoré buď medicínsky obnovujú (napr. neuralink implementovaný ochrnutému človeku) alebo umelo vylepšujú (augmented human) schopnosti človeka (**z pohľadu etiky ide o dva diametrálne odlišné spôsoby využitia**);
- víziu prepojenia plne funkčných pokročilých systémov AI a ľudského mozgu, čo by zahŕňalo problémy oscilujúce medzi výzvami spojenými s etikou uvedomelej umelej inteligencie a psychológiou, filozofickým a legislatívnym pohľadom na definíciu ľudskej osoby, morálnymi aspektami i vierou postmoderného človeka.

Javí sa, že aktuálny vývoj v oblasti moderných systémov AI stále viac zahŕňa problematiku zariadení z oblasti „skin-out“ technológií a implantovaných zariadení. Vylepšenia človeka na úrovni koexistencie s pokročilými systémami AI (napr. na úrovni BAGI/HLAGI) sú v súčasnosti skôr futurologickou záležitosťou.

Keďže pokročilé nástroje AI, nositeľné zariadenia a poniektoré zariadenia rozšírenej reality sú spoločensky široko akceptované a postupne sa stávajú predmetom bežného každodenného využívania, **pri masívnom prieniku týchto systémov AI do reálneho života treba rátať s možným vplyvom transhumanistickej filozofie**, ako napr. s mravným redukcionizmom a relativizmom, s falošnou premisou, že technické „vylepšovanie“ človeka v kontexte umelej inteligencie vedie k jeho kultúrnemu i morálnemu zlepšovaniu[314], materialistickému a technokratickému chápaniu šťastia a naplnenia,[315]

313 Ide o problémy súvisiace s vplyvom virtuálneho sveta na život človeka a spoločnosti, teda problémy týkajúce sa slabej umelej inteligencie, ktoré sú analogické problémom s ostatnými aspektami informačných technológií, kybernetického priestoru a virtuálneho sveta.
ŠANTAVÝ, *Analýza virtuálneho sveta v kontexte náboženskej formácie, pôsobenia a života Cirkvi,* [on-line], s. 25-47. [cit. 8. januára 2022]. Dostupné na internete: <https://peter.santavy.cloud/data/uploads/docs/analyza_virtualneho_sveta_v_kontexte_nabozenskej_formacie_posobenia_a_zivota_cirkvi.pdf>

314 STRAHOVNIK, V. *Virtues and transhumanist human enhancement*. In: PETROUŠEK R., ŽALEC B., eds. *Transhumanism as a Chalange for Ethics and Religion*. 2021, s. 37 – 44.

315 *Transhumanist Declaration*. [on-line]. [cit. 26. januára 2022]. Dostupné na internete: <https://hpluspedia.org/wiki/Transhumanist_Declaration>

s falošným rovnostárstvom s odovzdaním vlády superpočítačom (~umelej inteligencii)[316] a s pokriveným vnímaním eschatologickej roviny viery v kontexte transfigurizmu, t.j. náboženského transhumanizmu.[317]

1.12. Súčasnosť – umelá inteligencia „na koni“

V súčasnosti sme svedkami enormného rozmachu vo vývoji, tvorbe, realizácii, nasadení, poskytovaní i využívaní technológií umelej inteligencie. Tento rozmach stojí na viacerých pilieroch, ktoré sú ovocím súčasného rozvoja informačnej spoločnosti:

- pokrok vo vývoji metód a systémov umelej inteligencie, osobitne v oblasti neurónových sietí a generatívnych systémov;
- existencia, pomerne jednoduché získavanie a bezprecedentná dostupnosť[318] extrémneho množstva štruktúrovaných i neštruktúrovaných dát;[319]
- moderné prístupy a pokrok vo vede i v jednotlivých odvetviach, ktoré vyžadujú adaptívne procesy schopné spracovávať obrovské kvantá dát;
- potreba automatizácie – rozvoj automatizácie smerujúci k potrebe autonómnych systémov schopných samostatnej adaptívnej činnosti pri spracúvaní veľkého

316 LEE, I. *Equalism: Paradise Regained*. In: LEE N. (ed.). *The Transhumanist Handbook*. Springer, 2019, s. 849 – 863.

317 Solídny náhľad do celej tejto problematiky v optike kresťanského svetonázoru a Katolíckej cirkvi uvádza doc. Vivoda v svojom článku Transhumanizmus a Katolícka Cirkev.
VIVODA, M. *Transhumanizmus a Katolícka Cirkev*. In: *Nové Horizonty*. 2021, roč. 15, č. 3, s. 121-128. ISSN: 1337-6535, EAN 977133765400605

318 Už v roku 2010 Erich Smidt, v tom čase CEO Google, na konferencii Techonomy v kalifornskom Lake Tahoe vyhlásil: „V súčasnosti každé dva dni vyprodukujeme toľko informácií, koľko sme vytvorili od úsvitu civilizácie až do roku 2003.“
SIEGLER, M. G. *Eric Schmidt: Every 2 Days We Create As Much Information As We Did Up To 2003* [on-line]. [cit. 12. decembra 2021].
Dostupné na internete: <http://techcrunch.com/2010/08/04/schmidt-data/>

319 Senzorické systémy, počítačové siete, digitalizácia, informatizácia vedy i technologických procesov a predovšetkým on-line priestor komunikácie a sociálnych sietí sú zdrojom veľkého množstva dát. „Big data“, ako súčasť technologickej a informačnej revolúcie, umožnili rozvoj viacerých v súčasnosti dominantných metód umelej inteligencie, keďže vďaka veľkému množstvu dát je možné dostatočne a relevantne trénovať systémy AI, aby boli použiteľné v reálnom nasadení.
Por. MITCHELL, *Artificial Intelligence*, s. 80.

množstva neštruktúrovaných dát v konkrétnej oblasti;

- využívanie nástrojov umelej inteligencie s nástupom generatívnych systémov smerujúce k prirodzenému ľudskému spôsobu komunikácie a k aplikácii postupov, ktoré majú potenciál radikálne meniť obsah i spôsob našej práce;[320]
- možnosť integrovať nástroje AI do širokého spektra pracovných činností až po možnosť tímovej kooperácie a nahradenia ľudských zamestnancov;
- jednoduchá a lacná dostupnosť nástrojov AI pre bežnú populáciu a možnosti jej využitia v pracovnej, spoločenskej i osobnej sfére.

Tieto piliere sú súčasťou tvoriacej sa a (v niektorých častiach sveta) rozvíjajúcej sa informačnej a znalostnej spoločnosti, ktorá je sprevádzaná paradigmatickou zmenou: informačné technológie a dáta sa z roviny prostriedku dostávajú do roviny kontextu spoločnosti až do tej miery, že na základe zmien v technológiách a vzhľadom na prevratný vedecký i technologický pokrok sa mení medziľudská komunikácia v esenciálnej rovine, menia sa vzťahy, mení sa spoločnosť a princípy jej fungovania.[321]

Nástup a rozšírenie systémov umelej inteligencie patrí k tejto zmene paradigmy, pričom si dovolíme tvrdiť, že smerovanie k AGI a dosiahnutie systémov skutočnej umelej inteligencie by pre mnohých bolo jej naplnením.

320 Konverzačné rozhranie generatívnych systémov AI umožňuje komunikáciu v bežnom jazyku formou dialógu a pokročilej – akoby tímovej – spolupráce medzi človekom a strojom, pri ktorej systémy AI môžu vykonávať komplexné úlohy na pozadí tejto konverzácie.

321 ŠANTAVÝ, *Analýza virtuálneho sveta v kontexte náboženskej formácie, pôsobenia a života Cirkvi,* [online], s. 19-20. [cit. 7. decembra 2021].
Dostupné na internete:
<https://peter.santavy.cloud/data/uploads/docs/analyza_virtualneho_sveta_v_kontexte_nabozenskej_formacie_posobenia_a_zivota_cirkvi.pdf>