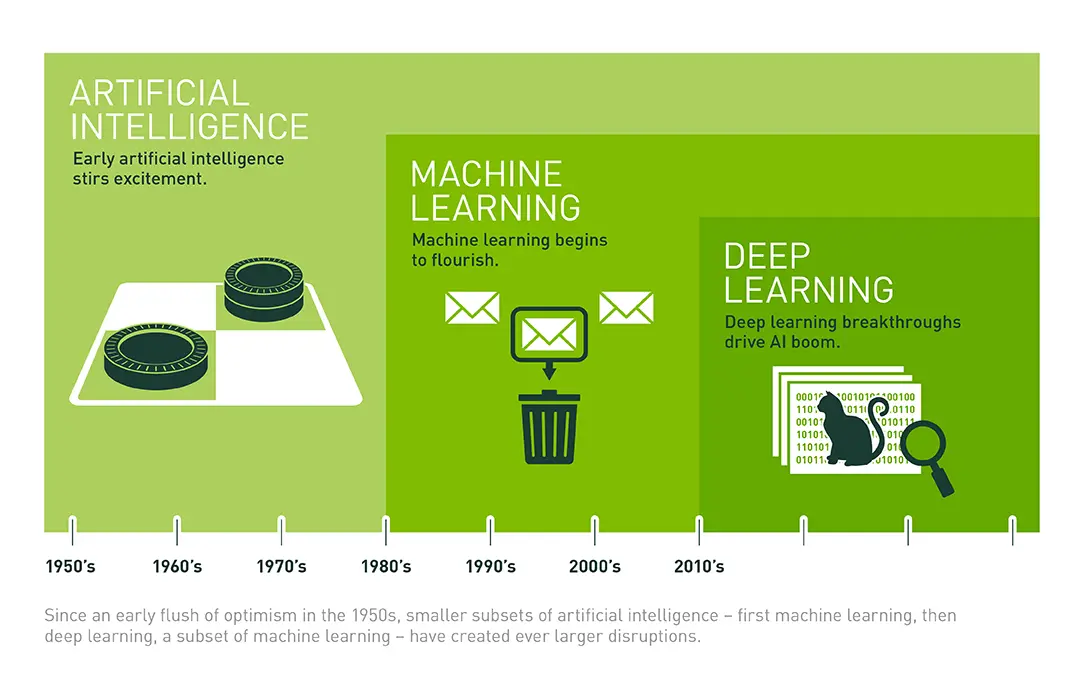
# AI, strojové a hluboké učení: jak se vlastně stroje "učí"?

3.7.2023, [Milan Šurkala](https://www.svethardware.cz/milan-surkala/autor/)

**Umělá inteligence** je něčím, co se probírá už roky. Teprve v poslední době se ale dostává i do povědomí lidí, kteří s technologiemi zrovna moc nekamarádí. Může za to především generativní AI **ChatGPT** od OpenAI, která přinesla v listopadu 2022 revoluci do světa IT. Máme tu chatbota pro každého, a to zdarma. Objevila se i spousta dalších systémů nejen pro generování textu, ale i generování obrazu, hudby a mnoha dalších věcí. Otázkou ale je, jak se takové stroje vlastně "učí"?



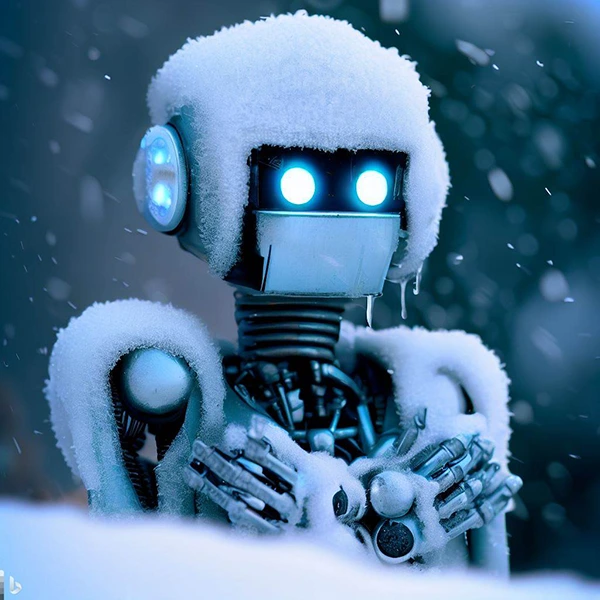
## Umělá inteligence

Aby to nebylo tak jednoduché, **umělou inteligenci** můžeme rozdělit na dvě velké oblasti, úzkou (Narrow AI, Weak AI) a obecnou (Strong AI, General AI). Pokud se tedy bavíme o AI, je dobré vědět, kterou část vlastně řešíme. Někomu můžete tvrdit, že využíváte AI, ale ten vám bude tvrdit, že o AI nejde, protože není univerzální jako člověk. Pravdu budete mít částečně oba. Jde totiž o to, o kterém typu AI se bavíte. Umělé inteligence jsou to v obou případech, ale jejich schopnosti jsou zásadně odlišné.

* **Narrow AI, Artificial Narrow Intelligence** (ANI) je AI, která je natrénována na specifickou činnost. To je většina AI systémů, se kterými se dnes setkáváme. Ta může rozpoznávat obrázky, řeč, optimalizovat fotografie, být základem asistentů jako Apple Siri, Amazon Alexa a mnoha dalších, může se také podílet na algoritmech v autonomních autech.
* **Strong AI, Artificial General Intelligence** (AGI) je obecnou AI, která by měla být ekvivalentem člověka. Neměla by být tedy schopna dělat jen jednu činnost, ale má být schopna dělat v podstatě vše a na vše se být schopna naučit. Také by měla být vědoma si sama sebe i plánovat budoucnost (což je něco, čeho se mnozí obávají). Ta je zatím spíše snem než realitou.

Kdy ale něco takového vzniklo? Pokud odhlédneme od slavné Čapkovy hry R.U.R. z roku 1921, která dala vzniknout slovu "robot", myšlenka umělé inteligence se datuje především do roku 1950, kdy **Alan Turing** vydal článek Computing Machinery and Intelligence, kde zároveň publikoval tzv. **Turingův test**. Ten říká, že stroj  vykazuje inteligentní chování v případě, že člověk není schopen při textové konverzaci se strojem i člověkem rozpoznat, kdy konverzuje s člověkem, a kdy se strojem. Nyní si jen v rychlosti prolétneme dosavadní historii událostí spojených s umělou inteligencí. V roce 1952 vznikla první aplikace schopná strojového učení od Arthura Samuela pro hraní dámy. Termín umělé inteligence jako takový pak vznikl v roce 1956 a jeho autorem je John McCarthy. A. Samuel se pak zasadil o vznik termínu "strojové učení" v roce 1959.

V roce 1966 vznikl první chatbot ELIZA a o rok později Frank Rosenblatt vytvořil první neuronovou síť postavenou na perceptronech, nicméně dalších cca 15 let se nic až tak technologicky revolučního nedělo a bylo i sníženo financování výzkumu AI (někdy označováno jako 1. AI-zima v druhé polovině 70. let), přestože v 80. letech už byly algoritmy využívající neuronové sítě a systém backprogration pro trénování docela běžné (spíše než technologicky a schopnostmi se to posunulo rozšířením). Rok 1984 se do učebnic AI "zapsal" spíše negativně, do mysli mnohých lidí zasel obavu z AI, která vyhladí lidstvo, legendárním filmem Terminátor. Zklamání z toho, že pořádná (obecná) AI stále nepřichází, vedlo ke 2. AI-zimě na přelomu 80. a 90. let, kdy bylo opět v mnoha zemích významně zkráceno financování výzkumů umělé inteligence.



AI-zima, vytvořeno pomocí generativní AI DALL-E v Microsoft Bing

 V roce 1997 však superpočítač **Deep Blue** od IBM porazil známého šachistu Garryho Kasparova. Měli jsme tu ale např. i první robotický vysavač Roomba z roku 2002 nebo dva rovery Spirit a Oportunity od NASA, které se po povrchu Marsu navigovaly autonomně bez pomoci člověka (2003-2004). Další porážka šampionů počítačem se stala v roce 2011, kdy **IBM Watson** porazil i vítěze hry Jeopardy! Kena Jenningse a Brada Rutta. V roce 2014 chatbot Eugene Goostman vyhrál Turingův test a rok 2016 přinesl (vedle pořádného průšvihu Microsoftu s chatbotem Tay) porážku světového šampiona ve hře Go. O to se postaral **AlphaGo** od DeepMind, který nyní patří pod Google. Pak už to šlo docela rychle, 2021 přišel DALL-E, a roky 2022 a 2023 přinesly především systém **ChatGPT** (a jeho konkurenci), chatboty pro širokou veřejnost.

## Strojové učení

## 

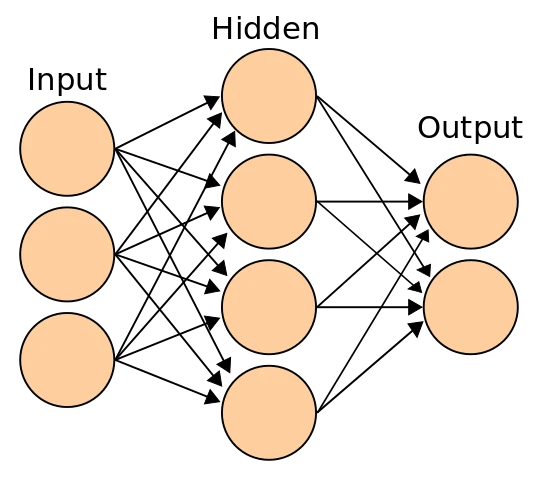
 Jak se učí roboti? (zdroj: generativní AI Adobe Firefly)

Dostáváme se k druhé části, a to **strojovému učení**. To už znamená, že stroji neříkáme explicitně to, co má dělat, ale nějakým způsobem ho necháme to naučit a nastavit si vlastní parametry. Předhodíme mu tedy velký počet trénovacích dat, aby se na nich naučil rozpoznávat a dělat to, co po něm požadujeme. Způsobů strojového učení je několik a každý má své vlastní poslání.

* **Učení s učitelem** (supervised learning) vyžaduje člověka, který algoritmu řekne, co má být výstupem (musí označit data - tohle se stane pracovní náplní mnoha z nás). Příkladem může být předhození milionů snímků koček a několika milionů snímků čehokoli jiného. Pokud u snímku s kočkou vyjde, že to kočka není, v případě využití algoritmů hlubokého učení, k čemuž se ještě dostaneme, to bude třetí část, viz zelený snímek od Nvidie výše, jsou parametry algoritmu nastaveny špatně a musí se upravit. Zpravidla to funguje tak, že se spočítají chyby, upraví parametry a zkusí se to znovu. Algoritmus se učí pomocí minimalizace chyby opakovanými pokusy tak dlouho, dokud tato celková chyba rozumně klesá. Je nutné si uvědomit, že změna parametru může pomoci některým případům, kde byl problém, ale zároveň může "rozbít" i případy, které už fungovaly. Navíc se to nesmí přehnat, protože by mohlo u některých algoritmů dojít k tzv. přetrénování. Tehdy by se dokonale naučil rozpoznávat případy z trénovací množiny, ale o to hůře by pak pracoval s jinými daty, které v ní nejsou (a které budou v reálném světě), což je něčím, co nechceme. Algoritmus má být univerzální a pracovat obecně. Učení s učitelem se využívá např. ke klasifikaci, detekci spamů, podvodů i předpovědím vývoje. Spadají sem mnohé algoritmy využívající neuronové sítě, ale i jiné jako např. SVM, Naive Bayes, různé regrese.
* **Učení bez učitele** (unsupervised learning) je něco, co na první pohled nedává smysl. Algoritmu neřeknete, co má vyjít. V tomto případě tak hledá skupiny vstupů, které mají nějaké společné vlastnosti. Nedokáže tedy říci, co daná data představují, ale dokáže říci, že patří do stejné skupiny. Toto se dá např. použít na streamovacích službách (tzv. recommendation systems, tedy systémy pro doporučování), kde se ale současně často používají i algoritmy z první skupiny. Když do systému hodíte uživatele, kteří sledují filmy a seriály (např. na Netflixu), jejich hodnocení filmů, to, na co se dívali, co se jim líbilo a co ne, dokážete identifikovat skupiny podobných uživatelů a pak na základě těchto informací o příslušností do skupin doporučovat. Učení bez učitele tedy hledá skryté vzory v datech, používá se často ke klastrování (hledání podobných objektů s podobnými vlastnostmi). Zajímavostí je, že systém na hledání koček z příkladu výše s učitelem lze udělat i učením bez učitele, jak se v roce 2012 povedlo dvěma výzkumníkům z Googlu (Jeff Dean and Andrew Ng) na 10 milionech videí z YouTube.
* **Zpětnovazební učení** (reinforcement learing) známé také jako učení posilováním. V tomto případě se systém snaží najít rovnováhu mezi průzkumem (exploration) a využití již naučeného (exploitation). Jinak řečeno, ví se, co vede ke správnému výsledku (za kterou získá "odměnu"), ale občas zkouší i nové věci pro případ, že by to vedlo k ještě lepšímu výsledku (k větší "odměně"). Využívá se v robotice, návrhu strategií (nechcete opustit fungující strategii, ale co když trochu jiná strategie bude ještě lepší?). Z příkladů lze zmínit např. Q-Learning nebo jeho verzi s neuronovými sítěmi Deep Q-Learning.

## Hluboké učení

Na AI i strojovém učení se pracuje už desítky let, základem moderní umělé inteligence a velmi účinného strojového učení se však staly až **neuronové sítě** (ty už také mají za sebou hodně dlouhou historii), což vedlo ke vzniku **hlubokého učení** (tyto dvě věci jsou někdy rozděleny na dvě různé vrstvy). To ale neznamená, že AI je postavena jen a pouze na neuronových sítích, nicméně byly to právě ony a jejich poslední průlomová vylepšení, která po roce 2010 systémy umělé inteligence pořádně "nakoply", ačkoli neuronové sítě jako takové existují už půlku století. Neuronové sítě existují v obrovském množství různých typů, snaží se napodobovat fungování lidského mozku a liší se mnoha více či méně zásadními detaily. V zásadě jde ale o to, co je u strojového učení obvyklé, tedy že máte nějaké vstupy (jednoduše data, která chcete vyhodnotit), a výstupy (zpracované výsledky z těchto dat, třeba výsledek, co má systém provést, do které skupiny má daný obrázek přiřadit,...). Rozdíl je v tom, že mezi tím je několik vrstev digitálních navzájem spojených neuronů, i proto se jim říká **ANN** (Artificial Neural Networks, umělé neuronové sítě).



Neuronová síť, autor: en:User:Cburnett, [*CC BY-SA 3.0*](http://creativecommons.org/licenses/by-sa/3.0/), přes Wikimedia Commons

Funguje to jednoduše tak, že každý umělý neuron si bere data ze všech neuronů předchozí vrstvy, každému z nich ale přiřazuje jinou váhu. Pokud bychom to měli třeba v autě v automatické převodovce, jedním vstupem (vstupní "input" vrstva) by mohl být právě zařazený stupeň, druhým aktuální otáčky a třetím např. zatížení motoru. První neuron (ve skryté "hidden" vrstvě) by dával těmto parametrům váhy např. 50 %, 30 % a 20 % (v podstatě ohodnocení vlivu jednotlivých šipek do něho vedoucích), zatímco druhý neuron by měl tyto váhy pro tatáž vstupní data jiné. Výsledné číslo (součet pronásobení všech vstupů příslušnými unikátními vahami) pak ovlivňuje chování tohoto neuronu pro to, co pošle do další vrstvy, v našem případě výstupní ("output"). Např. nedosáhne-li výsledek určité hodnoty, nic nepošle a podobně. A i zde, každá tato šipka má jinou váhu, jiné "procentní zastoupení". Výsledkem ve výstupní vrstvě může být to, zda převodovka nechá zařazený stupeň, přeřadí nahoru nebo naopak podřadí. Tréninkem mohou být miliardy, biliony jízdních situací všeho od mopedu až po 80tunový tahač a Formuli 1, přičemž se zkouší, zda pro danou jízdní situaci dává neuronová síť stejný výstup jako v trénovací situaci.

Neuronová síť je v podstatě jen docela "hloupý" systém pro nás úplně abstraktního sčítání násobených čísel a jejich přeposílání dál. Učení sítě je jen způsobem, jak najít taková čísla pro šipky, která pro dané vstupy budou po celém tomto kolečku mnoha sčítání a násobení dávat smysluplné výstupy. Tomuto učení se říká **trénování neuronové sítě**. Aby se síť natrénovala, potřebuje obrovské množství pokusů a obrovské množství dat v trénovací množině. Opět to funguje na principu minimalizace chyby, takže když neodpoví tak, jak by měla (v případě učení s učitelem), musí se upravit váhy, aby byl příště výsledek lepší. K tomu se využívá např. algoritmus back-propagation (zpětná propagace). Poté to zkusí znova a dělá to tak dlouho, dokud se chyba rozumně zmenšuje, což se nijak zásadně neliší od toho, jak to funguje u strojového učení obecně.

Myšlenkou je, že když bude dobře odpovídat na data z dostatečně velké a dostatečně obecné trénovací množiny (trénovacího datasetu) zahrnujícího co nejvíce možností, bude pravděpodobně dobře odpovídat i na data z reálného provozu. Průšvihem může být, když máte pro stejné vstupy dva rozdílné kýžené výstupy. Co se má pak naučit?

Verzí neuronových sítí je mnoho, liší se počtem vrstev, způsobem výpočtu, úprav hodnot a mnoha dalších detailů. Také se mohou lišit svým určením, které je ovlivněno tím, jak pracují. Např. konvoluční neuronové sítě (Convolutional Neural Network – CNN) se využívají zejména pro analýzu obrazových dat, generativní neuronové sítě (Generative Adversarial Network – GAN) míří naopak k vytváření nového obsahu na základě naučeného.



Mark Zuckerberg, autor: Anthony Quintano from Honolulu, HI, [*CC BY 2.0*](https://creativecommons.org/licenses/by/2.0), přes [*Wikimedia Commons*](https://commons.wikimedia.org/wiki/File:Mark_Zuckerberg_F8_2018_Keynote.jpg) (přidány otazníky)

Průšvih tu je ještě jeden. Tím, že se to vlastně učí stroj sám, a to i bez výraznější pomoci člověka, nikdo vlastně neví, jak to v daném konkrétním případě pořádně funguje. Příkladem může být Facebook a tlak různých regulátorů na to, aby zveřejnil své algoritmy. Tohle ale může být docela problém. Využívá-li se tu v různých místech AI, která může mít docela vágně definovaný cíl (např. prodloužit dobu strávenou na síti, zvýšení příjmů,...), ona sama sérií pokusů a omylů přijde na to, které parametry jak co ovlivňují. Samotný programátor v takovém případě nemusí být ten, kdo explicitně do kódu pro doporučování příspěvků "nadatloval", jak se to má dělat a komu co doporučovat. Algoritmus ke zveřejnění tak v podstatě ani pořádně nemusí existovat.

Pro vývoj algoritmů je potřeba extrémně výkonného hardwaru a mnohé systémy se trénují dny i týdny na tisících i desetitisících GPU (proces trvá velmi dlouho, i u jednoduchých sítí může jít o desítky tisíc pokusů), zatímco pro běh už stačí relativně normální hardware a poběží vám s trochou štěstí i v telefonu. Natrénovaný systém je zjednodušeně řečeno seznamem těchto parametrů, podle kterého už vyhodnotíte výsledek, když mu dodáte data z reálného provozu. Nalezení optimálních parametrů je tak obrovským problémem, jejich aplikace na danou situaci mnohem jednodušší. I tak to ale vyžaduje speciální enginy pro jejich běh (NPU) nebo výkonná GPU.

Docela podobný vztah je u těžby kryptoměn. Tam např. hledáte speciální číslo "nonce", které když se přidá k transakcím, má vytvořit hash určitého tvaru. Toto je výpočetně velmi složité a vyžaduje obrovský počet pokusů, než se takové číslo podaří najít. Jakmile se ale toto číslo najde a zveřejní, všem ostatním stačí jeden pokus k opětovnému vytvoření hashe s tímto nalezeným číslem pro ověření, že má opravdu ten tvar, který má mít. Zde se také provádí mnoho pokusů pro vyladění systému a nalezení čísel, na kterých bude systém AI fungovat, samotné fungování pro jednu konkrétní situaci je už jen jeden průchod sítí a získání výsledku (pochopitelně, v případě real-time systému toto děláte v podstatě neustále, protože je systém pořád v nových situacích).

Tato neuronová síť ale nemusí mít jen jednu skrytou vrstvu. Těch může být (a zpravidla také je) mnohem více a pak hovoříme o tzv. **hluboké neuronové síti**. Celé toto znamená, že nyní programátoři neřeší až tak moc programování samotného chování algoritmu, ale také to, jak se má správně učit. Samotné chování je dáno především poskytnutím dostatečně obřího množství dat, na kterém se naučí. I proto např. Musk tvrdí, že vývoj autonomních aut je především o najetí potřebného počtu miliard mil. To ale pochopitelně neznamená, že v autě běží jedna "neuronka", do které se jen nasype kvantum dopravních situací. Je zde komplexní (a v zásadě i docela tradiční) algoritmus pro řízení, kde mu různé systémy AI dopomáhají v rozpoznávání a rozhodování. Problémem je ale to, aby se to z toho obrovského množství dat rozumně naučilo, a zde je právě oblast, které se věnuje výzkum neuronových sítí. Stroj se sice může učit i jinak než pomocí neuronových sítí, ale právě ty jsou tím, co algoritmy pro roboty posunulo technologicky mnohem dále.

## Závěr

Jak vidíme, oblast umělé inteligence je hodně široká a to jsme to vzali jen po povrchu. Nyní už trochu lépe rozumíme tomu, co se v těch algoritmech děje, i tomu, proč mohou kvůli způsobu svého trénování vykazovat různé chyby. V příštím díle se více podíváme na systémy generativní AI, kam spadá např. již zmíněný ChatGPT nebo různé generátory obrazu jako Midjourney, DALL-E, nebo Adobe Firefly. Poslední dva se postaraly o generování obrázků do tohoto článku.

Zdroj: [ibm.com](https://www.ibm.com/topics/artificial-intelligence), [techtarget.com](https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence), [medium.com](https://arshren.medium.com/supervised-unsupervised-and-reinforcement-learning-245b59709f68), [zcu.cz](https://www.kiv.zcu.cz/studies/predmety/uzi/Folie_ZS/Stroj_uceni.pdf), [livescience.com](https://www.livescience.com/47544-history-of-a-i-artificial-intelligence-infographic.html), [washington.edu](https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf), [tableau.com](https://www.tableau.com/data-insights/ai/history), [javapoint.com](https://www.javatpoint.com/history-of-artificial-intelligence), [nvidia.com](https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/), [simplilearn.com](https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/ai-vs-machine-learning-vs-deep-learning)