IT a anatomie firmy

(AI, Umělá inteligence)

(pracovní dokument)

Mariia Shevchenko

VŠE Praha, 2026

Obsah

Účelem textu je ***základní teoretické vymezení umělé inteligence*** a jeho význam. Nejprve vymezuje samotný ***pojem umělé inteligence*** z různých perspektiv a stručně představuje její historický vývoj. Následně popisuje ***základní strukturu AI***, především strojové učení, neuronové sítě, hluboké učení a generativní umělou inteligenci včetně velkých jazykových modelů. Dále se zaměřuje na porovnání tradiční a generativní AI a ***na legislativní rámec*** upravující využívání těchto technologií.

1. Vymezení pojmu umělá inteligence

1.1 Definice umělé inteligence

Pojem umělá inteligence lze definovat z několika různých úhlů pohledu. Z pohledu počítačových věd je umělá inteligence ***obor, který se zabývá syntézou a analýzou výpočetních agentů***, kteří jednají inteligentně. Výpočetní ***agent j***e přitom vnímán jako i***nteligentní tehdy, pokud dokáže se flexibilně přizpůsobovat měnícímu se prostředí,*** dosahuje stanovených cílů, učí se ze svých zkušeností a činí vhodná rozhodnutí i přes omezené výpočetní kapacity (Poole & Mackworth, 2017).

V odborné literatuře se setkáváme i s dalšími definicemi. Stuart Russell a Peter Norvig ve své knize *Artificial Intelligence: A Modern Approach* systematicky rozdělili ***různé přístupy k definování*** umělé inteligence do čtyř základních kategorií. Tyto kategorie vycházejí ze dvou hledisek: ***myšlení vs. jednání a lidské vs. racionální***. Toto členění shrnuje Tabulka 1-1.

Tabulka 1-1: Definice umělé inteligence (Zdroj: Russell & Norvig, 2010)

Thinking Humanly	Thinking Rationally
"The exciting new effort to make computers think . . . machines with minds, in the full and literal sense." (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . ." (Bellman, 1978)	"The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)
Acting Humanly	**Acting Rationally**
"The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)	"Computational Intelligence is the study of the design of intelligent agents." (Poole et al., 1998) "AI . . . is concerned with intelligent behavior in artifacts." (Nilsson, 1998)

Z pohledu společenských věd je AI definována jako reálná ***schopnost nelidských entit vykonávat úkoly, řešit problémy, komunikovat*** a jednat logicky způsobem obdobným lidskému jednání. Tato definice kombinuje dvě klíčová hlediska: úroveň výkonnosti a úroveň autonomie (Gil de Zúñiga et al., 2023).

Na základě uvedených přístupů lze umělou inteligenci chápat jako ***souhrn metod, technik a systémů, které umožňují výpočetním agentům vnímat své okolí, analyzovat získaná data a učit se z nich***. Tyto systémy jsou schopny samostatně nebo částečně samostatně přijímat rozhodnutí a vykonávat činnosti směřující k dosažení předem stanovených cílů. Umělá inteligence nepředstavuje jednu konkrétní technologii, ale dynamicky se rozvíjející oblast, která zahrnuje různé přístupy. Společným

cílem těchto přístupu je napodobit nebo ***rozšířit vybrané aspekty lidské inteligence*** v digitálním prostředí.

1.2 Vývoj umělé inteligence

Historie umělé inteligence jako vědeckého oboru se ***začala v polovině 20. století***. V roce 1943 představili Warren McCulloch a Walter Pitts jednoduchý model formálního neuronu, čímž položili základy pro uvažování o umělých neuronových sítích. Termín „Artificial Intelligence“ poprvé systematicky použit v 50. letech 20. století profesorem Johnem McCarthym (Poole & Mackworth, 2017).

Za jeden z prvních zásadních momentů ve vývoji umělé inteligence je považována práce Alana Turinga z roku 1950. ***Turing navrhl koncept dnes známý jako Turingův test***, který posuzuje inteligenci stroje na základě jeho vnějšího chování. Pokud lidský hodnotitel při textové komunikaci nedokáže rozlišit, zda komunikuje s člověkem, nebo se strojem, lze podle tohoto kritéria stroj označit za inteligentní (Poole & Mackworth, 2017).

V tradičním období AI dominoval tzv. ***symbolický přístup***, založený na manipulaci symbolů a formálních pravidlech. Tento přístup vycházel z předpokladu, že ***inteligence*** může být realizována ***prostřednictvím logických reprezentací***. Typickým příkladem byly plánovací systémy využívající explicitní reprezentace stavů a akcí. Symbolická AI však čelila zásadním omezením, zejména vysoké výpočetní náročnosti a obtížím při aplikaci v reálném, dynamickém prostředí. To vedlo k rozvoji alternativních přístupů, zejména ***reaktivních architektur***, které nevyužívají explicitní model světa, ale reagují na podněty prostřednictvím pravidel typu podmínka – akce (Alonso, 2014).

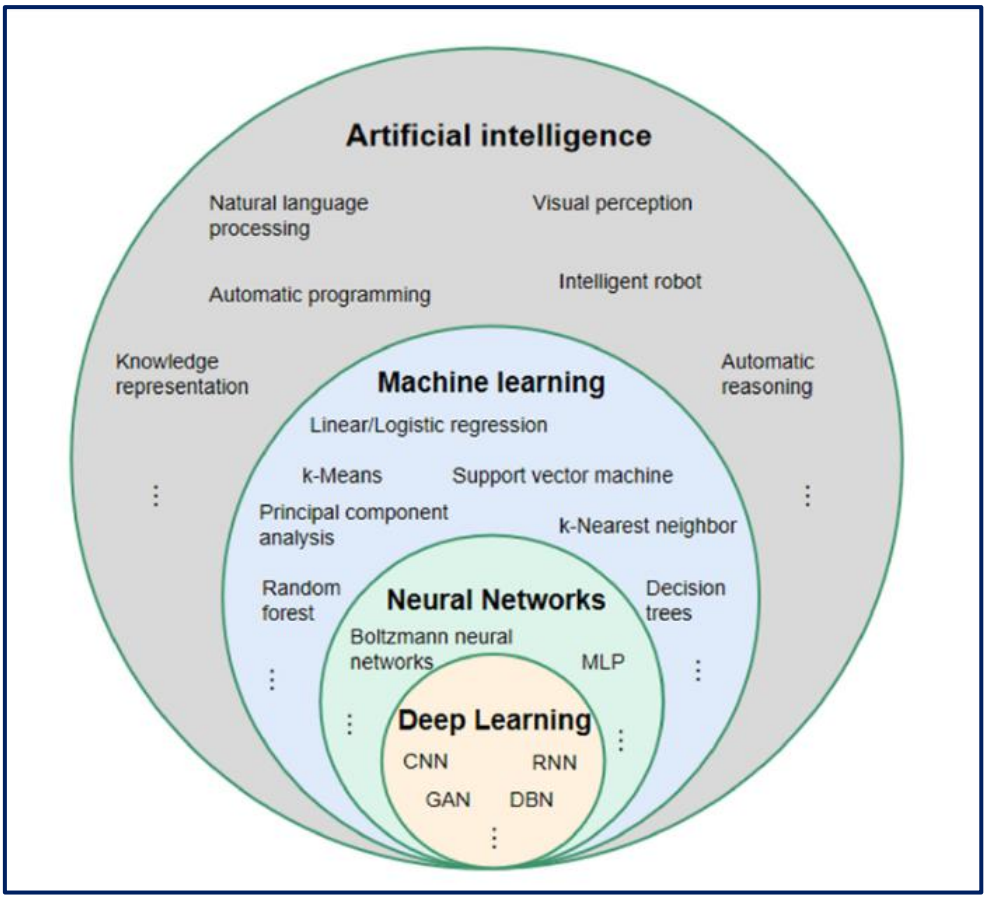
Od 90. let 20. století dochází k výraznému posunu směrem ***ke strojovému učení (Machine Learning),*** které umožňuje modelům učit se přímo z dat. Tento přístup je založen na statistických metodách a optimalizačních technikách a umožnil významné zlepšení výkonu v oblasti klasifikace, rozpoznávání obrazu či řeči. Následně se rozvinulo ***hluboké učení (Deep Learning),*** které využívá vícevrstvé neuronové sítě schopné automaticky extrahovat hierarchické reprezentace dat. Tyto modely dosáhly výjimečných výsledků v oblasti počítačového vidění, zpracování přirozeného jazyka a dalších aplikací (Hu & Hei, 2023).

Současná fáze vývoje umělé inteligence je charakterizována nástupem ***generativní umělé inteligence a velkých jazykových modelů***. Tyto modely, umožňují generovat komplexní textové výstupy a vykazují schopnost přizpůsobit se různým úlohám bez explicitního trénování pro každou z nich (Håkansson & Phillips-Wren, 2024).

Vývoj umělé inteligence tak lze charakterizovat jako postupný ***přechod od symbolických systémů přes statistické a datově orientované modely až po generativní a víceúčelové systémy***. Každá fáze ukazuje technologické možnosti své doby i měnící se požadavky společnosti.

1.3 Struktura umělé inteligence

Pro hlubší pochopení umělé inteligence je nezbytné vymezit její vnitřní strukturu a hierarchii. Oblast AI ***se skládá z mnoha subdisciplín***, které se do sebe hierarchicky zanořují. Obrázek 1-1 znázorňuje ***zjednodušenou strukturu umělé inteligence***.



Obrázek 1-1: Artificial intelligence (Zdroj: Li, 2021)

1.4 Strojové učení

Strojové učení je podmnožinou umělé inteligence zaměřenou na ***vývoj algoritmů, které se dokáží učit z dat bez nutnosti explicitního naprogramování všech pravidel***. Důraz je kladen na data-driven učení, kdy model na základě historických dat automaticky ***identifikuje vzory*** a využívá je pro predikci či rozhodování.

Základní ***dělení algoritmů strojového učení*** zahrnuje:

- Učení s učitelem (Supervised learning).
- Učení bez učitele (Unsupervised learning).
- Zpětnovazební učení (Reinforcement learning).

Učení s učitelem

Učení s učitelem představuje nejrozšířenější algoritmus strojového učení. Je založeno na práci ***s označenými daty, kde je pro každý vstup znám odpovídající správný výstup***. Cílem modelu je naučit se funkci, která mapuje vstupní proměnné na výstupní hodnotu tak, aby ***minimalizoval rozdíl mezi predikovaným a skutečným výstupem*** (Russell & Norvig, 2010). Proces učení obvykle probíhá optimalizací ztrátové funkce pomocí metod numerické optimalizace, například gradientního sestupu. Supervised learning se typicky ***dělí na klasifikaci a regresi***. V případě klasifikace je výstupem diskrétní třída, zatímco regrese pracuje s kontinuálními hodnotami. Modely využívané v rámci supervised learningu zahrnují ***lineární a logistickou regresi***, ***rozhodovací stromy, support vector machines i neuronové sítě*** (Goodfellow et al., 2016).

Praktické využití tohoto přístupu je velmi široké. V oblasti finančních služeb je supervised learning využíván například pro credit scoring, kde model na základě historických dat o klientech predikuje pravděpodobnost selhání (Purohit, 2023).

V textové analytice se používá ***pro klasifikaci dokumentů***, ***detekci spamu nebo analýzu sentimentu***. Výhodou tohoto přístupu je relativně vysoká přesnost při dostatečně kvalitních a reprezentativních datech. Nevýhodou je nutnost rozsáhlých označených datasetů. Tvorba takových datasetů může být nákladná a časově náročná (Russell & Norvig, 2010).

Učení bez učitele

Učení bez učitele ***pracuje s neoznačenými daty***, tedy ***bez předem definovaného správného výstupu***. Cílem není predikovat konkrétní hodnotu, ale ***odhalit v datech vnitřní struktury, vztahy či skryté vzory*** (Russell & Norvig, 2010). Model se snaží identifikovat podobnosti mezi pozorováními nebo nalézt latentní proměnné, které vysvětlují variabilitu dat. Typickými úlohami unsupervised learningu jsou ***shlukování a detekce anomálií***. Shlukování umožňuje segmentaci zákazníků podle podobnosti jejich chování, což je významné například v marketingu nebo řízení vztahů se zákazníky Detekce anomálií pomáhá při odhalování podvodných transakcí nebo neobvyklého chování v síťovém provozu. (Goodfellow et al., 2016).

Učení bez učitele je často využíván jako ***předzpracování dat pro další modelování*** nebo v situacích, kdy nejsou k dispozici označená data. Hlavní výhodou tohoto přístupu je nezávislost na ruční anotaci dat. Na druhé straně je interpretace výsledků často složitější než u supervised learningu (Russell & Norvig, 2010).

Zpětnovazební učení

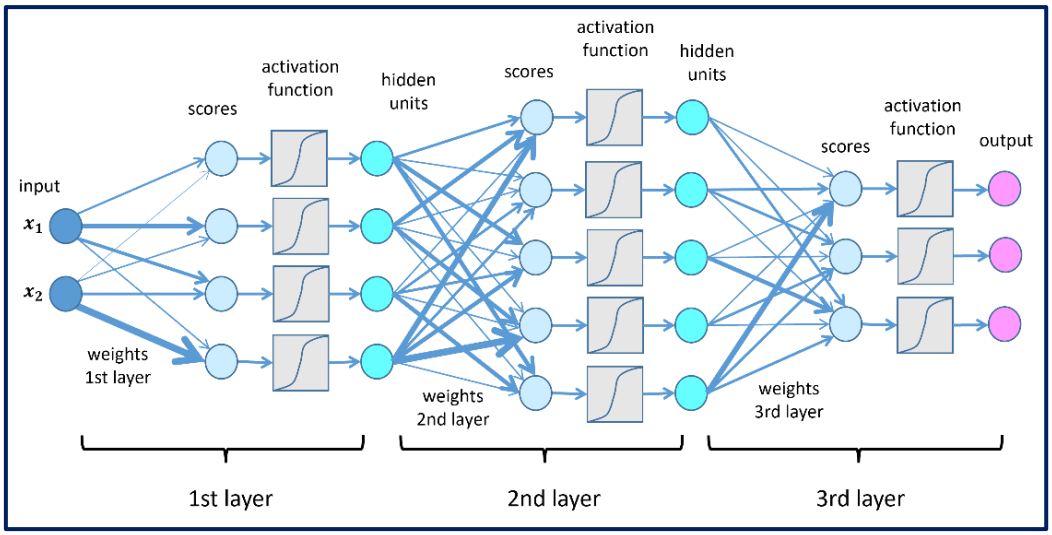
Zpětnovazební učení představuje specifický přístup, kdy se ***agent učí prostřednictvím interakce s prostředím***. Na rozdíl od učení s učitelem zde nejsou k dispozici explicitní správné výstupy pro jednotlivé kroky. Agent místo toho získává zpětnou vazbu ve formě odměn nebo penalizací a jeho cílem je najít takovou strategii, která maximalizuje celkový užitek v průběhu času. (Sutton & Barto, 2015).

Zpětnovazební učení nachází uplatnění zejména v oblastech, kde je ***rozhodování závislé na dlouhodobých důsledcích***, například v robotice, autonomních systémech nebo řízení portfolia. V posledních letech bylo zpětnovazební učení úspěšně aplikováno také v oblasti herních strategií a optimalizace komplexních procesů. Výhodou tohoto algoritmu je schopnost učit se optimální strategie i v dynamickém prostředí. Nevýhodou ale je vysoká výpočetní náročnost, potřeba velkého množství interakcí s prostředím a potenciální nestabilita trénovacího procesu (Sutton & Barto, 2015).

1.5 Neuronové sítě

Neuronové sítě představují ***klíčovou architektonickou složku*** umělé inteligence. Inspirací pro jejich vznik byla biologická struktura lidského mozku a způsob, jakým jsou neurony propojeny a předávají si informace. Umělé neuronové sítě jsou ***tvořeny vrstvami výpočetních neuronů, které zpracovávají vstupní signály pomocí vážených spojení a nelineárních aktivačních funkcí*** (Poole & Mackworth, 2017).

Základní strukturu neuronové sítě zobrazuje Obrázek 1-2 a zahrnuje vstupní vrstvu, jednu či více skrytých vrstev a výstupní vrstvu. Každý neuron přijímá vstupy, násobí je vahami, sčítá je a aplikuje aktivační funkci, která určuje výsledný výstup. ***Učení neuronové sítě probíhá úpravou vah*** na základě minimalizace ztrátové funkce, obvykle prostřednictvím algoritmu zpětného šíření chyby (back propagation) a gradientního sestupu. V kontextu strojového učení lze neuronové sítě chápat jako ***univerzální funkcí, schopné modelovat komplexní nelineární vztahy mezi vstupními a výstupními proměnnými***. Jejich výhodou je schopnost automatické extrakce příznaků, což snižuje potřebu ručního inženýrství vstupních charakteristik, které bylo typické pro tradiční modely strojového učení (Hu & Hei, 2023).



Obrázek 1-2: Structure: Deep neural networks (Lamarr Institute, 2021)

Rozvoj hlubokých neuronových sítí znamenal zásadní posun. ***Přidáním většího počtu skrytých vrstev dochází k hierarchickému zpracování dat.*** Takovým způsobem nižší vrstvy zachycují jednodušší vzory jako jsou hrany v obraze, zatímco vyšší vrstvy reprezentují abstraktnější koncepty jako objekty či významové struktury v textu. Tento princip je zásadní například pro ***konvoluční neuronové sítě*** (CNN) používané v počítačovém vidění či ***rekurentní neuronové sítě*** (RNN) určené pro sekvenční data (Poole & Mackworth, 2017).

Neuronové sítě tak představují ***technologické jádro současné AI***. Umožňují realizaci pokročilých systémů strojového učení, hlubokého učení, velkých jazykových modelů i generativní umělé inteligence. Neuronové sítě však nejsou bez omezení. Vyžadují ***rozsáhlé množství dat a vysoký výpočetní výkon,*** jsou náchylné k přeučení (overfitting) a jejich rozhodovací proces je často obtížně interpretovatelný. Otázka transparentnosti a vysvětlitelnosti neuronových modelů je proto významným tématem současného výzkumu (Hu & Hei, 2023).

1.6 Hluboké učení

Hluboké učení představuje specializovanou oblast strojového učení ***založenou na vícevrstvých neuronových sítích***, které umožňují hierarchické zpracování dat. Zatímco klasické neuronové sítě mohou obsahovat jednu nebo dvě skryté vrstvy, hluboké učení pracuje ***s architekturami obsahujícími desítky až stovky vrstev, což umožňuje modelovat komplexní nelineární vztahy*** mezi vstupy a výstupy. Rozvoj hlubokého učení byl umožněn zejména třemi faktory: dostupností rozsáhlých datových souborů (big data), zvýšením výpočetního výkonu a zdokonalením optimalizačních algoritmů. Kombinace těchto faktorů vedla ke zlepšení výkonu modelů v oblasti počítačového vidění, rozpoznávání řeči a zpracování přirozeného jazyka (Hu & Hei, 2023).

Jedním z klíčových přínosů hlubokého učení je ***schopnost automatické extrakce příznaků***. Zatímco tradiční ML vyžaduje ruční zadávaní vstupních charakteristik, hluboké neuronové sítě ***se učí relevantní reprezentace přímo z dat.*** V případě zpracování obrazu může první vrstva detekovat jednoduché rysy, jako jsou hrany, další vrstvy kombinují tyto rysy do složitějších tvarů a vyšší vrstvy identifikují celé objekty. Tento princip hierarchické reprezentace umožňuje modelům efektivně pracovat s velmi komplexními datovými strukturami (Purohit, 2023).

Hluboké učení je široce využíváno ***v oblasti autonomních vozidel***, kde neuronové sítě detekují objekty, jízdní pruhy a chodce, což umožňuje vozidlu reagovat na okolní prostředí. Dále je aplikováno při ***rozpoznávání obličejů pro účely biometrické autentizace*** a bezpečnostních systémů. V zemědělství umožňují analýzu satelitních snímků a senzorických dat za účelem optimalizace výnosů a řízení zdrojů. Významné uplatnění má taky ***v oblasti počítačového vidění***, ***zpracování přirozeného jazyka*** a rozpoznávání řeči. Přes své výhody je hluboké učení spojeno s řadou výzev. Modely vyžadují velké množství kvalitně anotovaných dat, jejichž příprava je časově i finančně náročná. Dalším problémem je ***omezené vysvětlení modelů***, které jsou často označovány jako „black box“ systémy (Purohit, 2023).

2. Generativní umělá inteligence

2.1 Principy fungování generativní AI

Generativní umělá inteligence představuje ***podkategorii umělé inteligence zaměřenou na tvorbu nového obsahu na základě naučených vzorů z*** rozsáhlých datových souborů. Na rozdíl od tradičních přístupů, které klasifikují, predikují nebo optimalizují, ***generativní modely vytváří text, obraz, zvuk, video nebo syntetická data***. GenAI využívá masivní korpusy dat získané z databází a webových zdrojů a generuje výstupy podobné na trénovací data. To znamená, že model netvoří obsah na základě porozumění, jak to dělá člověk, ale ***na základě statistických metod***. Z toho vyplývá i fenomén halucinací – generování věrohodně znějících, ale fakticky nesprávných informací (Håkansson & Phillips-Wren, 2024)

Z technologického hlediska jsou současné generativní modely ***založeny převážně na hlubokém učení a neuronových sítích s mnoha vrstvami***. Ve většině případů se jedná zejména o architekturu transformer, která umožňuje zachycovat vztahy mezi prvky v sekvencích dat prostřednictvím mechanismu self-attention (Chen et al., 2025)

Z širšího teoretického hlediska lze GenAI chápat jako posun od tradičního symbolického přístupu k datově řízeným modelům. Zatímco klasická AI pracovala s explicitními reprezentacemi a logickým odvozováním, současné ***generativní modely staví na distribuovaných reprezentacích vektorového prostoru***, kde význam vzniká ze statistických vztahů mezi prvky (Poole & Mackworth, 2017).

2.2 Large Language Models

Large Language Models jsou specifickou částí generativní AI zaměřenou na ***zpracování a generování přirozeného jazyka.*** LLM využívají ***hluboké neuronové sítě a rozsáhlé textové korpusy*** ke schopnosti rozpoznávat a generovat jazykové struktury.

Velké jazykové modely jsou ***charakterizovány počtem parametrů, škálou trénovacích dat a univerzálností použití.*** Jejich architektura umožňuje paralelní zpracování sekvencí a efektivní modelování dlouhodobých závislostí. Díky tomu modely dokáží generovat kontextově relevantní odpovědi, sumarizovat texty, překládat či odpovídat na otázky. LLM jsou trénovány pomocí self-supervised cílů, predikce následujícího slova (autoregresivní modelování) nebo doplňování maskovaných tokenů. Tímto způsobem si ***model osvojuje statistické reprezentace jazykových struktur*** bez explicitního anotování všech úloh. Výsledná schopnost generovat koherentní text je důsledkem ***zachycení pravděpodobnostních distribucí nad jazykovými sekvencemi***. (Chen et al., 2025).

Z hlediska autonomie lze LLM analyzovat pomocí dimenzí výkonu a autonomie. Umělá inteligence může vykonávat úkoly, ***činit rozhodnutí nebo vytvářet predikce***, přičemž míra autonomie závisí na rozsahu lidského zásahu. Velké jazykové modely typicky vykazují vysokou úroveň výkonu jako je generování textu nebo predikce odpovědí. Jejich autonomie ale je omezena vstupem uživatele a trénovacími daty. (Gil de Zúñiga et al., 2023)

LLM lze současně vnímat jako krok k obecnějším AI systémům. Koncept ***General-Purpose AI Systems (GPAIS) je schopen adaptace na více úloh bez explicitního přeprogramování.*** Generativní modely jsou v tomto rámci považovány za klíčový prvek širších, obecně použitelných AI systémů. (Triguero et al., 2023)

2.3 Prompt engineering

Prompt engineering představuje systematický ***návrh a optimalizaci vstupních pokynů (tzv. promptů),*** které usměrňují chování velkých jazykových modelů (LLM) tak, aby bylo dosaženo maximální přesnosti, relevance, soudržnosti a užitečnosti generovaného obsahu. Tato disciplína se postupně vyvinula ze stavu pokus-omyl do strukturované a formální výzkumné oblasti, která je klíčová pro plné využití potenciálu umělé inteligence. (Chen et al., 2025)

Návrh promptů lze rozdělit na ***základní techniky***, které formují strukturu a kontext zadání, a na ***pokročilé metodiky***, jež modifikují samotný způsob, jakým model uvažuje a zpracovává informace.

Podle Chen et al., 2025 ***k dosažení vysoce kvalitních výstupů*** z LLM je potřeba ***několik klíčových elementů v základním promptu:***

- ***Jasné a přesné instrukce***. Obecné pokyny vedou k tomu, že model čelí neomezenému množství možných interpretací, což ústí v příliš široké nebo vágní odpovědi. Jasné a specifické instrukce naopak zužují prostor pro odpověď a vedou k výstupům, které lépe odpovídají specifickým požadavkům dané situace.
- ***Role-based prompting***. Tato technika umožňuje modelu simulovat specifickou roli (např. "Jsi analytik"), čímž se generovaný výstup přizpůsobí požadovanému kontextu a odbornosti. Roli lze přiřadit buď staticky pro celou konverzaci, nebo dynamicky (role-play), kdy model upravuje své zaměření a výstupy na základě vyvíjejících ze vstupů uživatele v průběhu vícenásobných interakcí.
- ***Využití oddělovačů***. Běžné symboly (např. trojité uvozovky nebo speciální závorky) se používají k oddělení různých částí promptu. Pomáhají modelu přesně interpretovat strukturu vstupu a hrají kritickou roli v bezpečnosti – jasně oddělují uživatelská data od samotných instrukcí, čímž snižují riziko tzv. prompt injection útoků.
- ***Zero-shot, one-shot a few-shot prompting***. Tyto techniky definují, kolik příkladů model dostane před samotným řešením úlohy. Zatímco one-shot nebo few-shot poskytují modelu kontext a formu požadované odpovědi, zero-shot spoléhá výhradně na pretrénované znalosti modelu bez ukázek. Výzkumy navíc ukazují, že dobře navržený zero-shot prompt může v určitých scénářích překonat few-shot přístupy, protože příklady někdy mohou model spíš omezovat než učit.

Pro složitější analytické a logické operace byly vyvinuty ***pokročilé techniky, které strukturovaně vedou model k vyšší kvalitě výstupu.*** Mezi tyto techniky patří ***Chain-of-Thought, kde model ukazuje postupné kroky uvažování*** předtím, než vygeneruje finální odpověď, ***Tree of Thoughts, kde model zkoumá více možných strategií paralelně a Least-to-most,*** kde komplexní problém se rozděluje na jednotlivé jednodušší úlohy. (Chen et al., 2025)

Prompt engineering má taky kritický ***význam pro bezpečnost umělé inteligence***. Modely mohou být zranitelné vůči útokům typu prompt hacking, kdy útočník pomocí speciálně upraveného vstupu přinutí model ignorovat svá bezpečnostní omezení nebo vyzradit citlivé údaje. Právě proto je potřeba pečlivě strukturovat prompty a oddělovat datové části od instrukcí. (Chen et al., 2025)

2.4 Porovnání tradiční AI a generativní AI

Základní rozdíl mezi oběma přístupy ***spočívá v jejich hlavním cíli***. ***Tradiční*** AI je primárně navržena k řešení jedné specifikované úlohy, jejímž cílem je obvykle ***extrakce informací***, ***analýza dat nebo predikce.*** Modely tradiční AI přijímají vstupy a mapují je na specifické výstupy, jako jsou číselné hodnoty nebo kategorie. Typickými příklady jsou algoritmy jako rozhodovací stromy, logistická regrese nebo klasické konvoluční sítě pro rozpoznávání obrazu. Výstupem je tedy obvykle rozhodnutí, štítek nebo pravděpodobnostní hodnota. (Hu & Hei, 2023).

Na druhé straně ***GenAI slouží k tvorbě zcela nového obsahu***. Generativní modely se učí základní distribuci a vzory v trénovacích datech, což jim umožňuje ***generovat syntetické vzorky***, které jsou podobné původním datům, ale jsou originální. Dokážou vytvářet ***souvislé texty, fotorealistické obrázky, zvuk, videa či syntetická data*** (Håkansson & Phillips-Wren, 2024).

Dalším výrazným rozdílem je ***přístup k učení a nároky na trénovací data***. Tradiční systémy často spoléhají na učení s učitelem, které vyžaduje rozsáhlé sady pečlivě anotovaných a oštítkovaných dat. Příprava takových datových sad je časově i finančně náročná a omezuje schopnost modelu přizpůsobit se úlohám, pro které štítky neexistují. (Russell & Norvig, 2010). Naproti tomu ***generativní umělá inteligence využívá masivní objemy neoznačených dat,*** jako jsou texty z internetu, o velikosti gigabytů až terabytů. Spoléhají na self-supervised learning, kdy si model sám vytváří trénovací signály, jako je predikce skrytého slova ve větě nebo předvídání následujícího prvku v sekvenci. Tento postup modelu umožňuje absorbovat obrovské množství obecných znalostí o světě ještě předtím, než je vůbec aplikován na konkrétní úkol. (Triguero et al., 2023)

Schopnost generalizace je jednou z klíčových odlišností. ***Tradiční AI funguje v uzavřeném světě***. Předpokládá se, že všechny úkoly a třídy objektů, se kterými se systém setká, byly definovány již ve fázi tréninku. Pokud takový model narazí na novou, neviděnou třídu nebo odlišný typ problému, obvykle selže a pro jeho přizpůsobení je nezbytné systém zcela přetrénovat na nových datech. Když u ***GenAI se očekává, že budou schopny zvládat i takové úkoly, pro které nebyly výslovně a***

záměrně trénovány, a flexibilně se přizpůsobovat dynamicky se měnícímu prostředí. Díky mechanismům učení s malým počtem příkladů nebo vůbec bez příkladů lze model aplikovat na zcela novou doménu jen poskytnutím kontextu v zadání, aniž by se musely měnit váhy a parametry samotné neuronové sítě. (Triguero et al., 2023).

3. Legislativní rámec AI

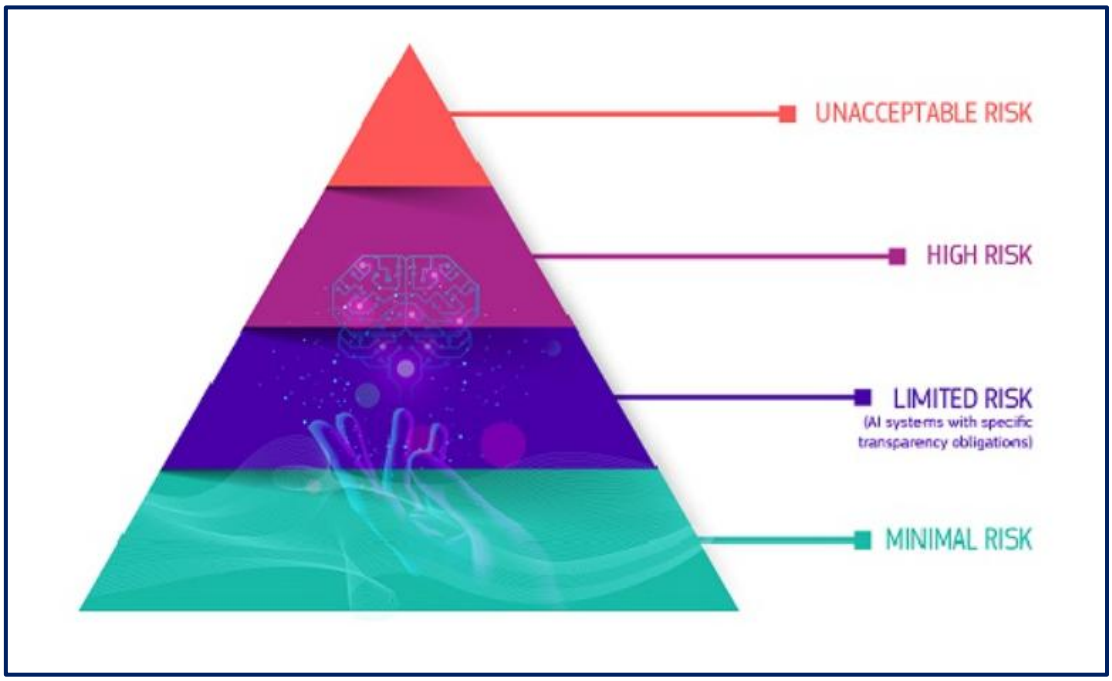
S rostoucím rozšířením umělé inteligence se objevují také ***potenciální rizika, jako je narušení soukromí nebo automatizované rozhodování bez dostatečné lidské kontroly***. Z těchto důvodů začaly státy a mezinárodní organizace vytvářet regulatorní rámce, jejichž cílem je zajistit bezpečné využívání umělé inteligence (Arda, 2024).

Regulace AI se obvykle zaměřuje ***na několik klíčových oblastí***. Mezi této oblasti patří ***ochrana základních práv jednotlivců, odpovědnost za škody*** způsobené autonomními systémy a ***zajištění bezpečnosti*** technologických řešení. Vzhledem k tomu, že umělá inteligence může zásadně ovlivňovat rozhodovací procesy v různých oblastech společnosti, je důležité vytvořit právní rámec, který zajistí rovnováhu mezi podporou inovací a ochranou veřejného zájmu (Sousa e Silva, 2024).

Jedním z nejvýznamnějších regulatorních kroků v oblasti umělé inteligence je ***přijetí Evropského aktu o umělé inteligenci (AI Act).*** Tento právní předpis představuje první komplexní legislativní rámec pro regulaci umělé inteligence na světě a jeho cílem je vytvořit bezpečné a důvěryhodné prostředí pro vývoj a využívání AI v Evropské unii. AI Act stanovuje harmonizovaná pravidla pro vývoj, uvádění na trh a používání systémů umělé inteligence v rámci EU. Hlavním cílem tohoto nařízení je minimalizovat rizika spojená s používáním AI a současně podporovat inovace a technologický rozvoj. ***Regulace je založena na principu důvěryhodné AI,*** která musí respektovat základní práva, bezpečnost a transparentnost (European Commission, 2024).

Klíčovým principem této legislativy je ***risk-based approach***, což je přístup založený na hodnocení rizika. AI systémy jsou ***klasifikovány podle míry rizika***, které mohou představovat pro jednotlivce nebo společnost. Tento model umožňuje přizpůsobit regulatorní požadavky podle potenciálního dopadu konkrétní technologie (European Commission, 2024).

AI Act rozděluje systémy umělé inteligence do několika kategorií podle jejich rizikovosti, jak ukazuje Obrázek 3-1. Každá z těchto kategorií podléhá odlišné míře regulace a kontrolních mechanismů.



Obrázek 3-1: A Risk-based Approach (Zdroj: European Commission, 2024)

Nejvyšší kategorii představují systémy ***s tzv. nepřijatelným rizikem***. Tyto systémy jsou považovány za neslučitelné s hodnotami Evropské unie a jejich používání je zakázáno. Patří sem například

technologie, které ***manipulují lidským chováním***, zneužívají zranitelné skupiny obyvatel nebo umožňují sociální hodnocení občanů (European Commission, 2024).

Další kategorií jsou ***vysoce rizikové systémy***, které mohou významně ***ovlivňovat zdraví, bezpečnost nebo základní práva jednotlivců.*** Mezi typické příklady patří AI používaná v kritické infrastruktuře, zdravotnických systémech, dopravě, vzdělávání nebo při rozhodování v oblasti zaměstnání. Tyto systémy musí splňovat přísné regulatorní požadavky, například zavedení systému řízení rizik, kvalitní správu dat, dokumentaci a lidský dohled nad rozhodovacími procesy (European Commission, 2024).

AI systémy ***s omezeným rizikem*** podléhají zejména ***požadavkům na transparentnost***. Typickým příkladem jsou chatboty nebo generativní systémy, u nichž musí být uživatel informován o tom, že komunikuje s umělou inteligencí. Transparentnost je důležitá zejména pro prevenci manipulace a zajištění informovanosti uživatelů (European Commission, 2024).

Poslední kategorií jsou systémy ***s minimálním nebo zanedbatelným rizikem***. Do této skupiny spadá většina běžných aplikací AI, například algoritmy doporučování nebo filtry nevyžádané pošty. Tyto systémy ***nepodléhají regulatorním požadavkům***, protože jejich potenciální dopad na společnost je relativně nízký (European Commission, 2024).

Legislativa týkající se umělé inteligence úzce souvisí s ochranou osobních údajů. V evropském právním prostředí hraje ***klíčovou roli zejména General Data Protection Regulation (GDPR),*** které upravuje zpracování osobních dat. Zatímco GDPR se zaměřuje především na ochranu osobních údajů jednotlivců, AI Act se soustředí na regulaci samotných AI systémů a jejich bezpečné používání. Tyto dva právní předpisy ***se vzájemně doplňují*** – GDPR zajišťuje ochranu dat používaných při trénování modelů umělé inteligence, zatímco AI Act stanovuje požadavky na transparentnost, bezpečnost a odpovědnost těchto systémů (Nolte, Rateike & Finck, 2025).

Mezinárodní organizace se taky snaží vytvářet rámce pro odpovědné využívání AI. Například Organisation for Economic Cooperation and Development publikovaly etické principy pro vývoj a využívání umělé inteligence, které zdůrazňují transparentnost, odpovědnost a ochranu lidských práv (OECD, 2019). Podobně UNESCO přijalo v roce 2021 dokument Recommendation on the Ethics of Artificial Intelligence, který představuje první globální normativní rámec pro etické využívání AI. (UNESCO, 2021)

Kromě mezinárodních organizací přijímají ***vlastní regulační přístupy také jednotlivé státy***. Například Spojené státy americké přistupují k regulaci AI spíše prostřednictvím doporučení, a standardů než jednotného zákona. Mezi jejich iniciativy patří například AI Bill of Rights, který definuje základní principy ochrany uživatelů, včetně lidského dohledu nad automatizovaným rozhodováním. Velká Británie zvolila flexibilní přístup založený na sektorové regulaci. Základní principy bezpečnosti, spravedlnosti a odpovědnosti jsou aplikovány prostřednictvím existujících regulačních institucí v jednotlivých odvětvích. Tento přístup byl představen v dokumentu A pro-innovation approach to AI regulation, jehož cílem je podpořit inovace a zároveň zajistit bezpečné využívání AI (AI Ethics and Integrity International Association, 2025).

Evropská unie je však považována za globálního lídra v oblasti regulace AI. ***Evropský AI Act je často vnímán jako model***, který může ovlivnit budoucí legislativu v dalších regionech světa a přispět k vytvoření mezinárodních standardů pro bezpečné využívání umělé inteligence.

4. Závěr

- Umělá inteligence představuje ***široký a dynamicky se rozvíjející obor***, jehož význam v současné společnosti i podnikové praxi neustále roste.
- AI se postupně vyvinula do podoby ***datově orientovaných a generativních systémů***, které jsou schopny nejen analyzovat informace, ale také vytvářet nový obsah a podporovat rozhodování v komplexních situacích.
- Velký význam v současnosti nabývá ***generativní umělá inteligence a velké jazykové modely***, které rozšiřují možnosti využití AI v praxi.
- Jejich ***přínos*** však není dán pouze technologií, ale také způsobem jejich používání, kvalitou vstupních dat a vhodným nastavením interakce prostřednictvím promptů.
- ***V kontextu risk managementu*** se umělá inteligence používá jako významný nástroj, který může organizacím pomoci přejít k prediktivnímu řízení rizik.
- AI ***umožňuje efektivnější identifikaci rizik, přesnější analýzu velkých objemů dat i automatizaci vybraných procesů.***
- Pro efektivní využití AI v oblasti řízení rizik je proto nezbytné chápat tuto technologii ***nikoli jako náhradu lidského úsudku***, ale jako nástroj, který jej může významně rozšířit a podpořit.

5. Zdroje

COSO (2004). *Enterprise Risk Management – Integrated Framework. Application Techniques.* Committee of Sponsoring Organizations of the Treadway Commission. Dostupné z: https://www.macs.hw.ac.uk/~andrewc/erm2/reading/ERM%20-%20COSO%20Application%20Techniques.pdf

COSO (2017). *Enterprise Risk Management – Integrating with Strategy and Performance.* Committee of Sponsoring Organizations of the Treadway Commission. Dostupné z: https://static.poder360.com.br/2023/09/Diretriz-Enterprise-Risk-Management-Coso-2017.pdf

Hill, T. & Westbrook, R. (1997). *SWOT Analysis: It's Time for a Product Recall. Long Range Planning*, 46–52. Dostupné z: https://ftms.edu.my/images/Document/MOD001074%20-%20Strategic%20Management%20Analysis/WK6_SR_MOD001074_Hill_Westbrook_1997.pdf

ISO (2018). ISO 31000. *International standard. Risk management — Guidelines*. Dostupné z: https://www.ler.uam.mx/Calidad-UAML/wp-content/uploads/2025/02/ISO-31000-2018.pdf

Majumder, S. & Dey, J. (2024). *A Notion of Enterprise Risk Management: Enhancing Strategies and Wellbeing Programs Available*. Emerald Publishing. ISBN: 978-1-83797-736-9. Dostupné z: https://doi.org/10.1108/9781837977352,

Rejda, G. et al. (2016). *Principles of Risk Management and Insurance*. Pearson Education. Boston. ISBN 978-0-321-41493-9, Dostupné z: https://fab.upt.edu.vn/wp-content/uploads/2020/04/Principles-of-risk-management-and-insurance_2016.pdf

Lamarr Institute (2021). *Deep Learning: How do deep neural networks work?* Dostupné z: https://lamarr-institute.org/blog/deep-neural-networks/

Li, S. et al. (2021). *A Comprehensive Review on Radiomics and Deep Learning for Nasopharyngeal Carcinoma Imaging*, 11(9), 1523. Dostupné z: https://doi.org/10.3390/diagnostics11091523

Smejkal, V. & Rais, K. (2013). *Řízení rizik ve firmách a jiných organizacích.* Praha: Grada Publishing a.s. ISBN 978-80-247-4644-9

Hopkin, P. (2018). *Fundamentals of Risk Management*. New York: Kogan Page. SBN 978-0-7494-8307-4

Hampton, J. (2009). *Fundamentals of Enterprise Risk Management : How Top Companies Assess Risk, Manage Exposure, and Seize Opportunity*. AMACOM. Dostupné z: http://ebookcentral.proquest.com/lib/vsep/detail.action?docID=452609

Acebes, F., González-Varona, J.M., López-Paredes, A. et al. (2024). *Beyond probability-impact matrices in project risk management: A quantitative methodology for risk prioritisation*. Humanit Soc Sci Commun 11. Dostupné z: https://doi.org/10.1057/s41599-024-03180-5

Gartner, (2025). *Magic Quadrant for Governance, Risk and Compliance (GRC) Tools, Assurance Leaders*. Dostupné z: https://www.logicgate.com/resources/reports/gartner-magic-quadrant-for-grc-tools-assurance-leaders/

Poole, D. L., & Mackworth, A. K. (2017). *Artificial Intelligence: Foundations of Computational Agents (2nd ed.)*. Cambridge University Press. Dostupné z: https://api.pageplace.de/preview/DT0400.9781108173773_A30911680/preview-9781108173773_A30911680.pdf

Alonso, E. (2014). *The Cambridge Handbook of Artificial Intelligence.* (pp. 232-246). UK: Cambridge University Press. ISBN 9781139046855. Dostupné z: https://openaccess.city.ac.uk/id/eprint/5191/1/CUP-RCO.pdf

Gil de Zúñiga et al. (2023). A Scholarly Definition of Artificial Intelligence (AI): Advancing AI as a Conceptual Framework in Communication Research. Dostupné z: https://www.scopus.com/pages/publications/85179989692?origin=resultslist#

Russell, S. J., & Norvig, P. (2010). *Artificial intelligence: A modern approach (3rd ed.).* Pearson. Dostupné z: https://people.engr.tamu.edu/guni/csce625/slides/AI.pdf

Hu, F., & Hei, X. (2023). *AI, machine learning and deep learning: A security perspective*. CRC Press. Dostupné z: https://www.scribd.com/document/840304414/AI-Machine-Learning-and-Deep-Learning-a-Security-Perspective-Fei-Hu-Z-Library

Håkansson, A., & Phillips-Wren, G. (2024*). Generative AI and large language models – Benefits, drawbacks, future and recommendations.* Procedia Computer Science. Dostupné z: https://doi.org/10.1016/j.procs.2024.09.689

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. Dostupné z: https://ai-kosh.indiaai.gov.in/static/Deep+Learning+Ian+Goodfellow.pdf

Purohit, A. (2023). *AI, ML, DL, and generative AI face off: A comparative analysis*. https://synoptek.com/insights/it-blogs/data-insights/ai-ml-dl-and-generative-ai-face-off-a-comparative-analysis/

Sutton, R. S., & Barto, A. G. (2015). *Reinforcement Learning: An Introduction*. Dostupné z: https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf

Chen, B., Zhang, Z., Langrené, N., & Zhu, S. (2025). *Unleashing the potential of prompt engineering for large language models*. Patterns. Dostupné z: https://doi.org/10.1016/j.patter.2025.101260

Triguero, I. et al. (2024). General purpose artificial intelligence systems (GPAIS): Properties, definition, taxonomy, open challenges and implications. Dostupné z: https://www.sciencedirect.com/science/article/abs/pii/S1566253523004517

Arda, S. (2023). *Taxonomy to regulation: A (geo)political taxonomy for AI risks and regulatory measures in the EU AI Act*. Dostupné z: https://www.semanticscholar.org/paper/Taxonomy-to-Regulation%3A-A-(Geo)Political-Taxonomy-Arda/c17befe2beb989ea8884b6a0a920839ed27b44b1

Sousa e Silva, N. (2024). *The Artificial Intelligence Act: Critical overview*. Dostupné z: https://www.nsousaesilva.pt/images/_Data/Publicacoes-outrosmateriais/EN_NSS_AI_Act.pdf

European Commission (2024). *Regulatory framework for artificial intelligence*. Dostupné z: https://commission.europa.eu/topics/business-and-industry/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en

European Commission (2024). *AI Act*. Dostupné z: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Farmer, D. (2025). *AI in risk management: Top benefits and challenges explained*. TreeHive Strategy. Dostupné z: https://www.techtarget.com/searchsecurity/tip/The-benefits-of-using-AI-in-risk-management

Wai, F. (2026). *AI in risk management: The executive guide to opportunities, challenges & use cases.* Dostupné z: https://thedigitalprojectmanager.com/project-management/ai-in-risk-management/

Yazdi, M., Zarei, E., Adumene, S., & Beheshti, A. (2024*). Navigating the power of artificial intelligence in risk management: A comparative analysis. Safety*, 10(2), 42. Dostupné z: https://doi.org/10.3390/safety10020042

Gurung, A. (2024). *Transformative ability of artificial intelligence in risk management (Bachelor's thesis)*. Arcada University of Applied Sciences. Dostupné z: https://www.theseus.fi/bitstream/handle/10024/859021/Gurung_Ambika.pdf?sequence=2

KPMG. (2026). *AI is revolutionizing risk management*. Dostupné z: https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2025/ai-revolutionizing-risk-management.pdf

Nolte, H., Rateike, M., & Finck, M. (2025). *Robustness and cybersecurity in the EU AI Act*. Dostupné z: https://blog.genlaw.org/pdfs/genlaw_icml2024/4.pdf

OECD. (2019). *OECD principles on artificial intelligence*. Organisation for Economic Co-operation and Development. Dostupné z: https://www.oecd.org/en/topics/ai-principles.html

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence. Dostupné z:* https://unesdoc.unesco.org/ark:/48223/pf0000380455

AI Ethics and Integrity International Association. (2025). *Global approaches to artificial intelligence regulation. Dostupné z*: https://ai-ei.org/global-approaches-to-ai-regulation-a-comparative-guideline/