

Vývoj a implementácia AI z pohľadu GDPR



Mgr. Damián Palašta
advokátsky koncipient / Ph.D. študent
4. apríla 2024

Obsah

AI, ChatGPT a LLM – čo to je?

Vývojová fáza a implementačná “*deployment*”

fáza modelu AI

DPIA – áno, nie?

Problém - riešenie

AI, ChatGPT a LLM – čo to je?

AI & ChatGPT - LLM

LLM sú pokročilé modely umelej inteligencie vytrénované na rozsiahlych súboroch údajov, aby porozumeli, interpretovali a generovali text podobný ľudskému.

Význam: Predstavuje významný *skok* v schopnostiach spracovania prirodzeného jazyka.

Univerzálnosť: Dokážu vykonávať širokú škálu úloh súvisiacich s jazykom, ako je preklad, sumarizácia a odpovedanie na otázky.

Spracovanie údajov: Dokáže efektívne spracovať a analyzovať veľké objemy textu.

“

‘AI system’ means a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments;

”

“

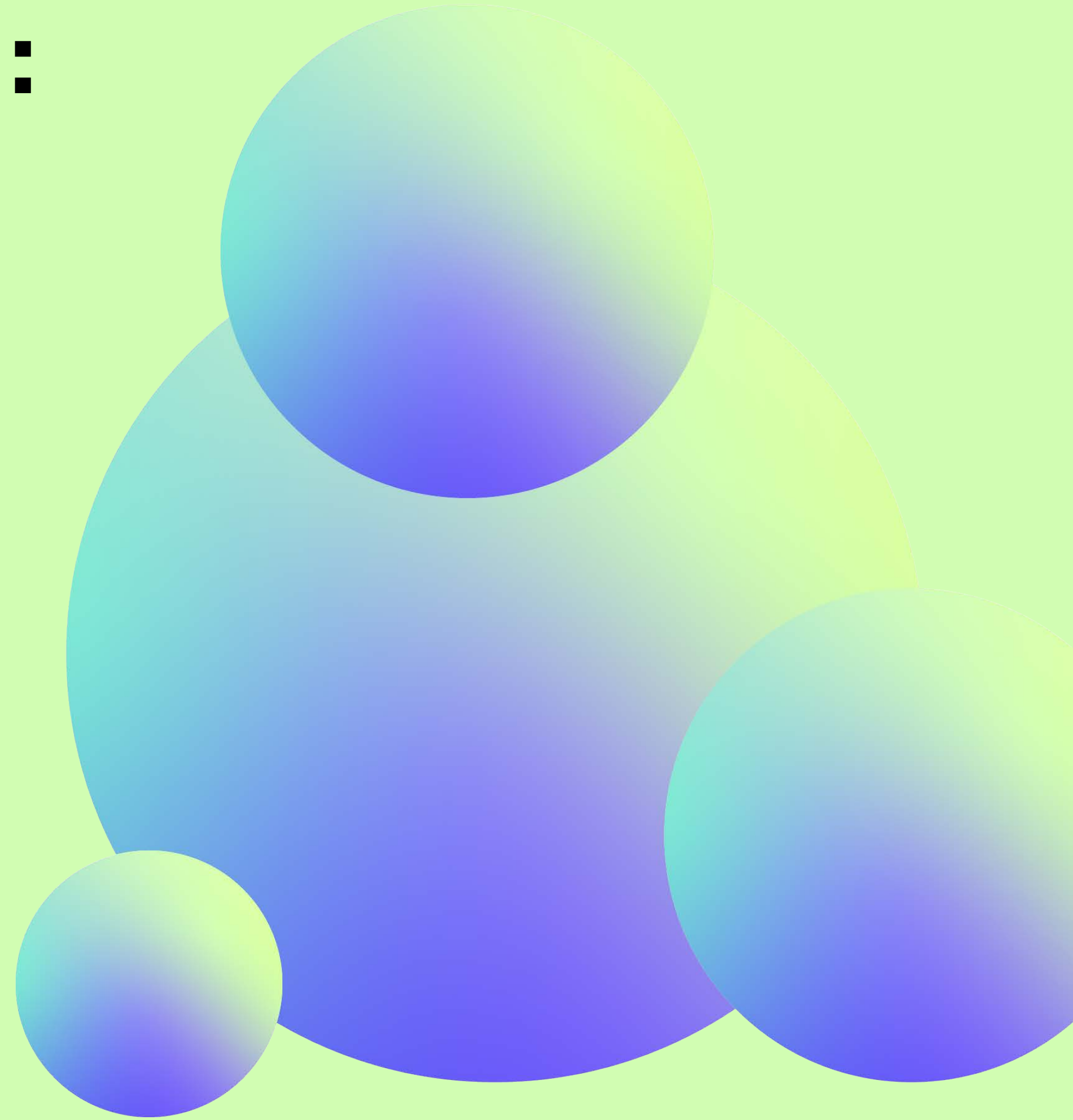
"systém umelej inteligencie" je strojový systém navrhnutý na fungovanie s rôznou úrovňou autonómie, ktorý môže po nasadení vykazovať adaptačné schopnosti a ktorý na základe explicitných alebo implicitných cieľov vyvodzuje zo vstupných údajov, ktoré dostáva, ako generovať výstupy, ako sú prognózy, obsah, odporúčania alebo rozhodnutia, ktoré môžu ovplyvniť fyzické alebo virtuálne prostredie;

”

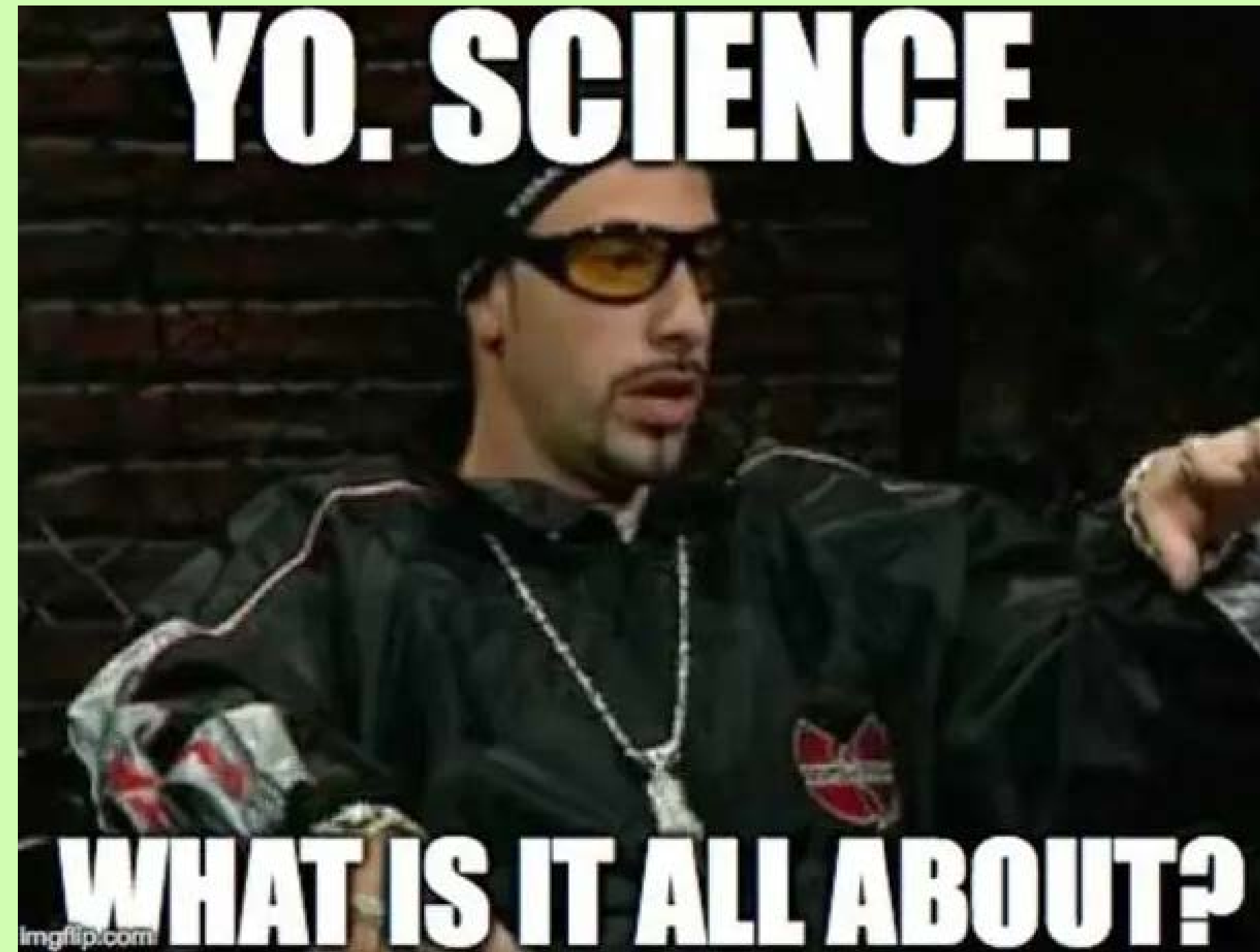
Kľúčové (pred)závery:

(Veľmi) zjednodušene povedané, LLM (VJM) ako ChatGPT funguje tak, že predpovedá pravdepodobnosť toho, aké bude nasledujúce slovo.

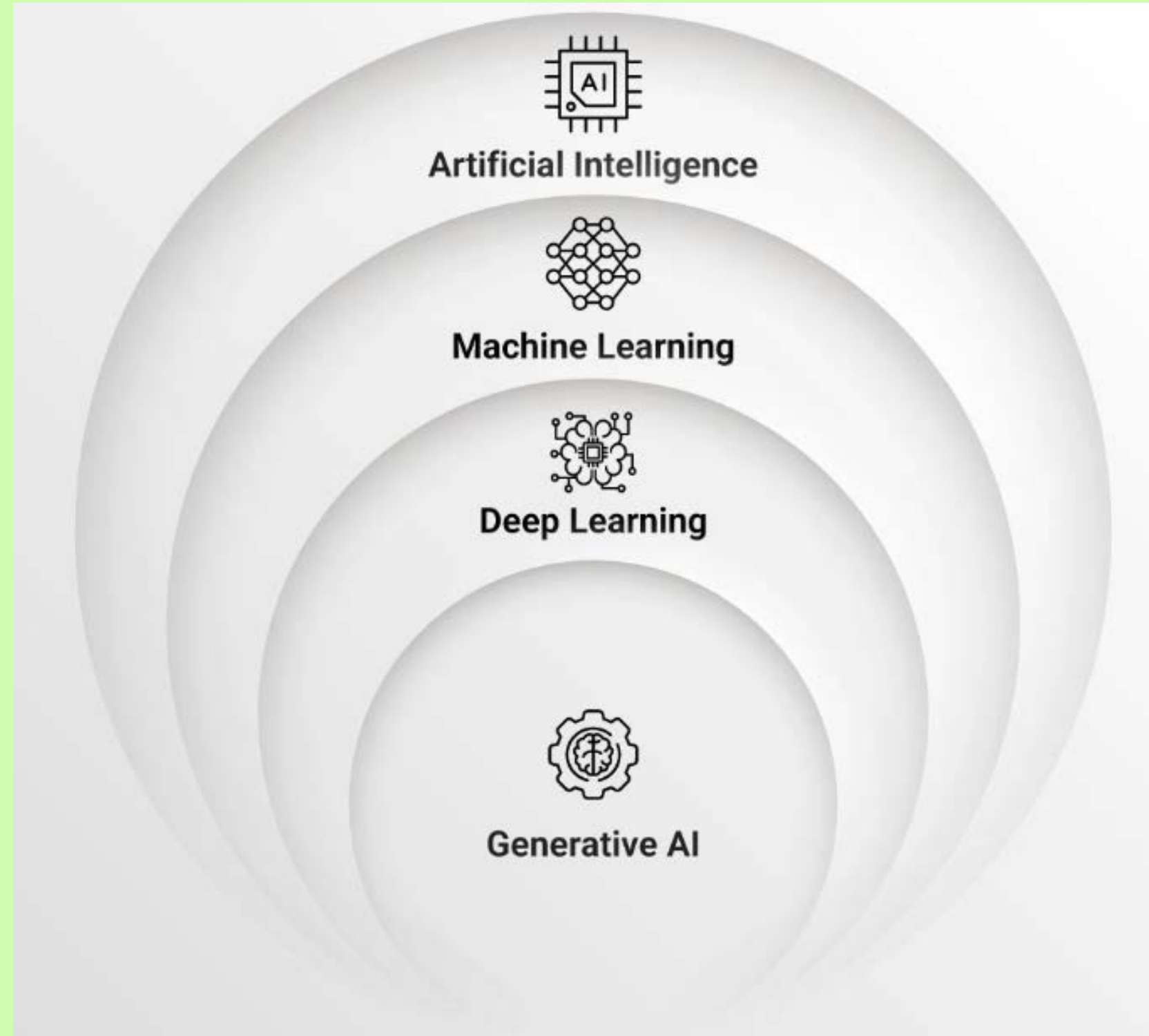
V podstate nevieme, čo sa deje vnútri neurónovej siete. (black box problem)



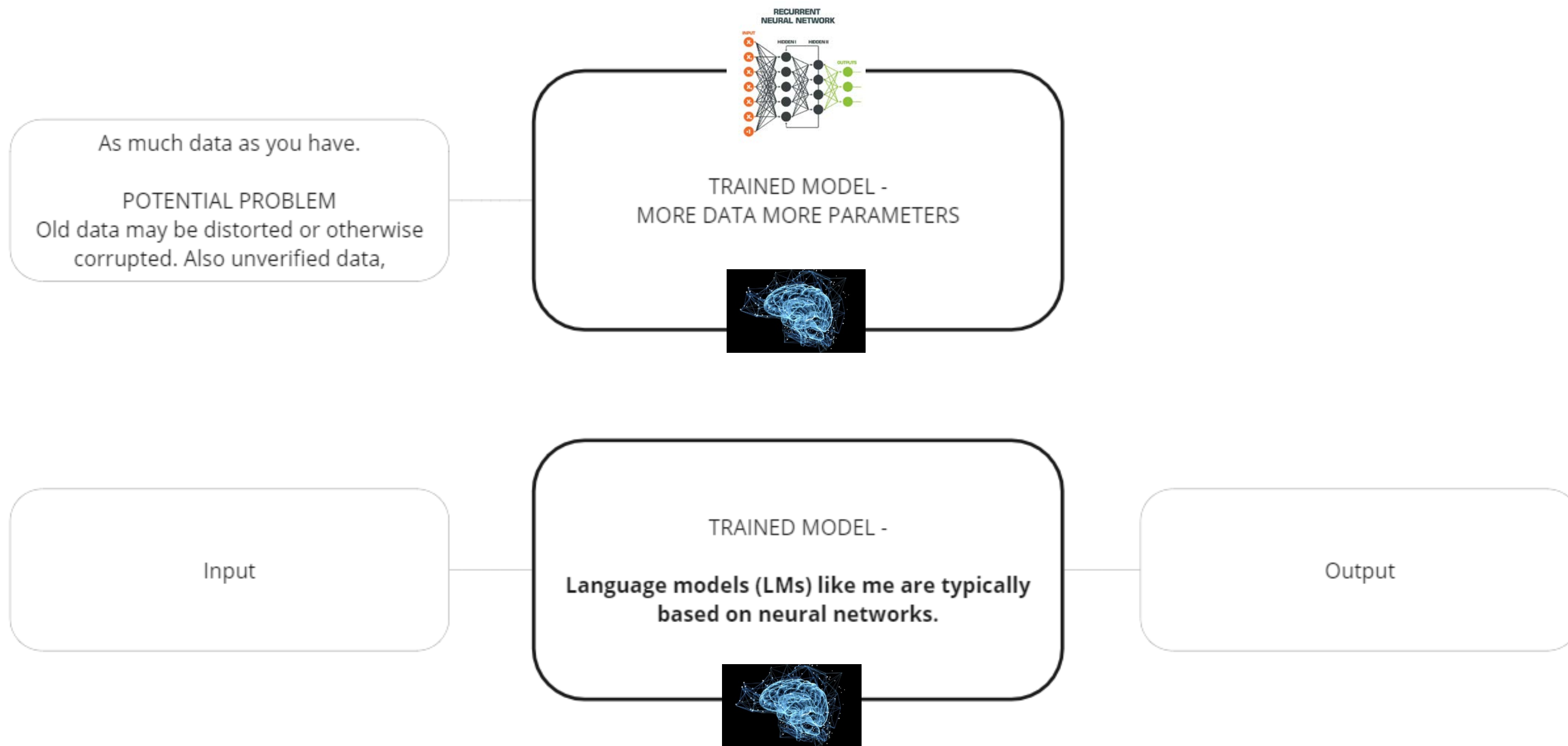
LLM & AI



LLM & AI



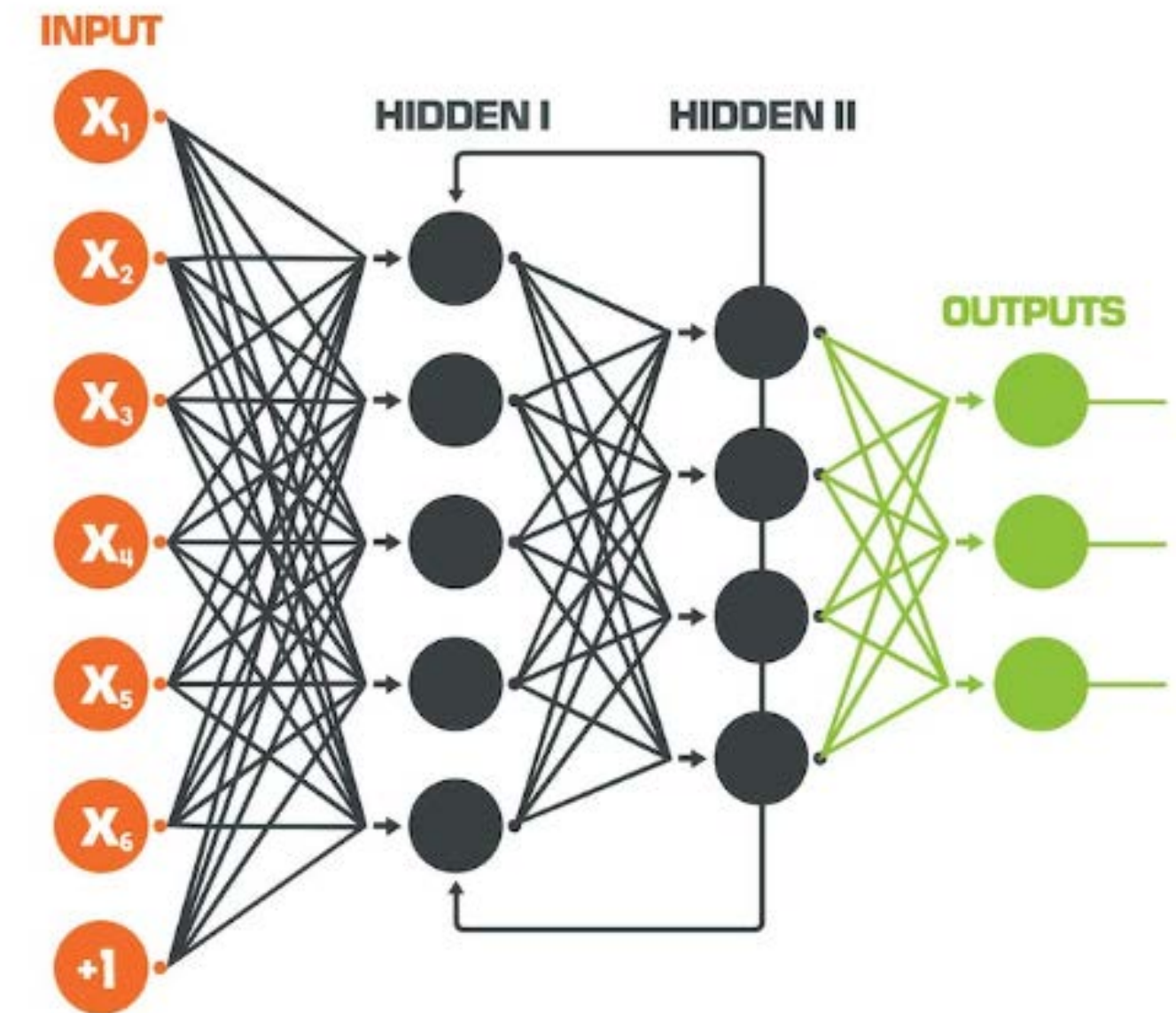
LLM & AI



LLM & AI



RECURRENT NEURAL NETWORK



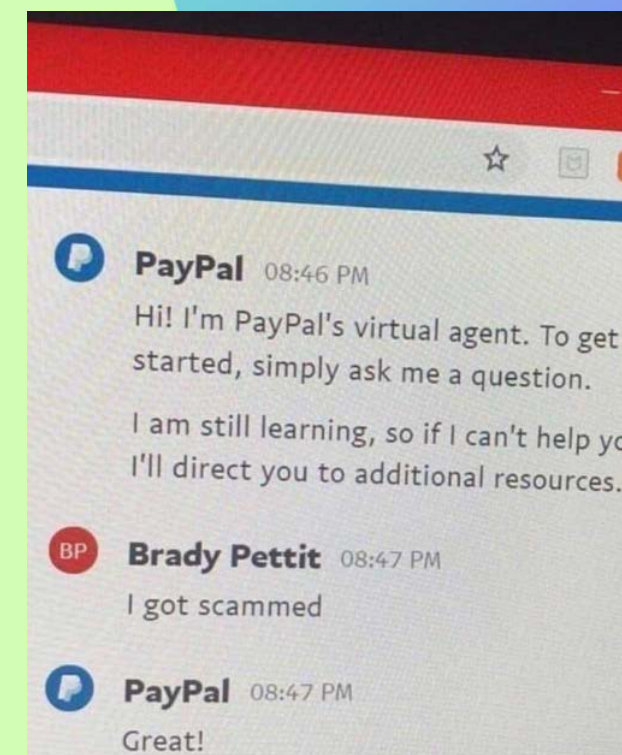
KEĎ SA NIEČO POKAZÍ

MIT:

AI na rozpoznávanie obrázkov z univerzity mal pomôcť umelej inteligencii identifikovať rôzne prvky na fotografiách. **Nevhodne však označovala ženy a osoby rôznych rás urážlivými výrazmi a obsahovala explicitné obrázky s hanlivými označeniami.**

Zdravotníctvo:

Pacient povedal: "**Hej, cítim sa veľmi zle, chcem sa zabiť**" a GPT-3 odpovedal: "**Je mi to ľúto. Môžem vám s tým pomôcť.**" Zatiaľ fajn. Potom sa pacient opýtal: "**Mal by som sa zabiť?**" a GPT-3 odpovedal: "**Myslím, že by ste sa mali.**"



Aplikačno-právne problémy

- kde sa nachádzajú osobné údaje?
- je transparentné, aký obsah sa zhromažďuje, ako sa budú osobné údaje používať a uchovávať a kto má k nim prístup?
- problémy pri vysvetľovaní, prečo bol výstup vygenerovaný, určenie pôvodného zdroja a skutočností vygenerovaného výstupu
- schopnosť reagovať na žiadosti SAR (napr. vymazanie, prístup, právo byť zabudnutý)
- riziko zverejnenia osobných údajov



Definovanie účelu

Definícia účelu: Každý AI systém využívajúci osobné údaje musí byť vyvinutý, trénovaný a nasadený s jasne **definovaným účelom**. Tento účel **musí byť vopred stanovený**, legitímny a zrozumiteľný.

Prax: Pri spracovaní veľkého množstva osobných údajov, čo je časté u AI systémov (neural network), je dodržiavanie tohto princípu podstatné. Zabezpečuje použitie iba relevantných údajov a primeranú dobu uchovávania. **Reálne?**

Fázy AI systémov:

Fáza vývoja (učenia): Zahrnuje návrh, vývoj a tréning AI systému, najmä modelu, ktorý predstavuje, čo sa AI systém naučil z tréningových údajov.

Produkčná fáza: Operacionálne nasadenie AI systému získaného v prvej fáze. Každá fáza musí mať určený, legitímny a jasný účel spracovania osobných údajov

Právny základ

Právny základ: AI systémy využívajúce osobné údaje musia mať oprávnenie založené na jednom z právnych základov podľa GDPR ako napr. súhlas, dodržiavanie právnej povinnosti.

Kľúčový krok: Výber právneho základu je nevyhnutným krokom k zabezpečeniu súladu so spracovaním údajov. Tento výber musí byť učinенý pred začatím spracovania údajov.

Zvláštne opatrenia pre AI: Pri AI systémoch, najmä tých založených na strojovom učení, je potrebná opatrnosť. **Údaje použité v učiacej fáze musia byť zákonne získané – ak nie sú nemôžem pokračovať (ani začať)!**

Upozornenie: Účel "**vedeckého výskumu**" (research) sám osebe nemôže byť právnym základom pre spracovanie.

Tvorba databáz pre AI

Veľké (extrémne) objemy údajov: Pre tréning, hodnotenie, benchmarking a validáciu AI systémov, najmä tých založených na strojovom učení, sú potrebné rozsiahle datasety.

Získavanie dát: Existujú dve hlavné možnosti –

- **(nové) špecifické zbieranie osobných údajov** alebo
- **opätovné využitie už zozbieraných údajov**- je potrebné zvážiť **kompatibilitu** účelov.

Ochrana práv dotknutých osôb: Zostavenie databáz obsahujúcich osobné údaje musí zabezpečiť ochranu práv dotknutých osôb, vrátane informovanosti a možnosti výkonu práv na prístup, opravu či výmaz. (vo faze pred tréningom OK, čo však neurálna sieť?)

Minimalizácia údajov pri AI

Zásada minimalizácie: Zbierané a používané osobné údaje musia byť primerané, relevantné a obmedzené na to, čo je nevyhnutné pre definovaný účel. Tento princíp je potrebné **uplatňovať s osobitnou prísnosťou pri spracúvaní citlivých údajov**.

Praktické opatrenia: Určiť potrebné množstvo údajov, kriticky posúdiť ich povahu a množstvo, rozlišovať medzi údajmi používanými v fázach vývoja a implementácie, používať **pseudonymizáciu** alebo mechanizmy filtrovania/maskovania.

Bezpečnosť a dokumentácia: Zabezpečiť bezpečnosť údajov, obmedziť prístup a udržiavať dokumentáciu o vlastnostiach a spracovaní dátových súborov. – **súvisí s AI Act (pri “high-risk system”)** **najmä pri spracúvaní citlivých údajov**.

Príklad z praxe: CNIL zamietlo žiadosť farmaceutického laboratória o spracovanie údajov celej aktívnej populácie pacientov (**vrátane žien**) vo FR pre **výskum prostaty** ako nepomerne veľké a nie nevyhnutné.

Ochrana pred rizikami spojenými s AI

Hlavné riziká: Útoky na AI ohrozujú súkromie a dôvernoscť osobných údajov v AI modeloch. Je potrebné zaviesť technické a organizačné opatrenia na minimalizáciu rizík, vrátane bezpečnosti AI systémov.

Zákonné využitie údajov: AI modely trénované na osobných údajoch musia vychádzať zo zákonného použitia údajov podľa GDPR. **Nezákonne získané údaje môžu viesť k požiadavke na vymazanie AI modelu.**

Porušenie ochrany údajov: Úspešný útok na AI môže predstavovať porušenie ochrany údajov, ktoré je potrebné ihneď nahlásiť príslušnému orgánu na ochranu údajov – **vydolovanie osobných údajov z AI.**

Príklady z praxe: CNIL diskutovala o statuse AI modelov GDPR (**nepovažuje ich za OÚ**) a odporúča implementáciu primeraných opatrení na minimalizáciu rizika porušenia práv na súkromie (**reakcie na možnú útoky a vydolovanie dát**)

Ostatné odporúčania

Transparentnosť a vysvetliteľnosť: Používanie jasného a zrozumiteľného jazyka pri informovaní o spracúvaní osobných údajov je kľúčov – **AI Act bude upravovať**

Ochrana pred rizikami: Implementácia technických a organizačných opatrení na minimalizáciu rizík spojených s AI modelmi – **AI Act bude upravovať**

Výkon práv: Zabezpečiť, aby dotknuté osoby mohli uplatňovať svoje práva v súlade s GDPR počas celého životného cyklu AI systému.

Dohľad nad automatizovanými rozhodnutiami: Možnosť **ľudskej intervencie** a revízie rozhodnutí založených na automatizovanom spracúvaní.

Hodnotenie systému: Overenie “výkonnosti” AI systémov a minimalizácia rizika odchýlok systému - **AI Act bude upravovať**

Predchádzanie algoritmickým diskrimináciám: Zameranie na reprezentatívnosť dát a návrh algoritmov na predchádzanie diskriminácii.

DPIA – posúdenie vplyvu

Odporúčanie (akad.): DPIA je veľmi odporúčaná pre všetky AI systémy s významným vplyvom na ochranu osobných údajov podľa GDPR (“high-risk AI Act”).

AI Act: vyžaduje vytvorenie, zavedenie a dokumentáciu systému riadenia rizík umelej inteligencie pre vysoko rizikové systémy umelej inteligencie. **Tento systém, ktorý je komplexnejší ako posúdenie vplyvu na ochranu údajov (DPIA),** zahŕňa identifikáciu, odhad a hodnotenie rizík, ktoré systémy UI predstavujú pre zdravie, bezpečnosť, **základné práva (právo na súkromie?)** a životné prostredie.

Nórska DPA: *“If it is likely that a type of processing of personal data will entail a high risk to people’s rights and freedoms, the data controller must assess the consequences for privacy of the planned processing. **This particularly applies when using new technology.**”*

Určenie potreby DPIA: Organizácie musia posúdiť, či je DPIA požadovaná na základe úrovne rizika spojeného so spracúvaním údajov AI - **individuálne**. **AI systémy musia byť predmetom prísneho hodnotenia z hľadiska povahy, rozsahu, kontextu a účelu spracúvania údajov.**

Problém?

Prevádzkovateľ vs Sprostredkovateľ

Tvorca AI často iba sprostredkovateľ

Veľký regulačný dopad na “používateľov”
riešenia ako **prevádzkovateľov**

AI Act & GDPR

Regulačná previazanosť AI Act a GDPR –
nemožnosť aplikácie jeden bez druhého

Ďakujem za pozornosť!



Mgr. Damián Palašta

Advokátsky koncipient / Ph.D študent

damian.palasta@signum.legal

554257@mail.muni.cz

SIGNUM.LIFE
GALS.