

Základy regulácie kybernetickej bezpečnosti a súvisiacich oblastí

Modul 5 - Širšie regulačné súvislosti kybernetickej bezpečnosti
v kontexte digitálnych politík Európskej únie

Časť. 2

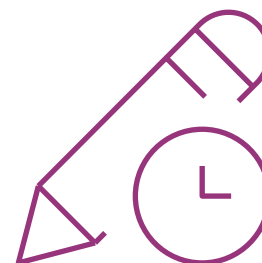
doc. JUDr. Matúš Mesarčík, PhD., LL.M

Kompetenčné centrum pre reguláciu kybernetickej bezpečnosti, ochrany súkromia a kybernetickej kriminality CUSEC



PRÁVNICKÁ FAKULTA
Univerzita Komenského
v Bratislave

Financované Európskou úniou Next GenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu pod číslom 17R05-04-VO1-00002



Agenda

- Akt o umelej inteligencii – pôsobnosť, základné pojmy, kľúčové inštitúty
- Akt o umelej inteligencii – prepojenie s kybernetickou bezpečnosťou
- Právne nástroje boja proti dezinformáciám v online priestore – rozmer práva EÚ
- Právne nástroje boja proti dezinformáciám v online priestore – rozmer slovenského právneho poriadku

REGULÁCIA UMELEJ INTELIGENCIE



OECD: Principles of AI

OSN: Global Digital Compact



Rada Európy: Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law

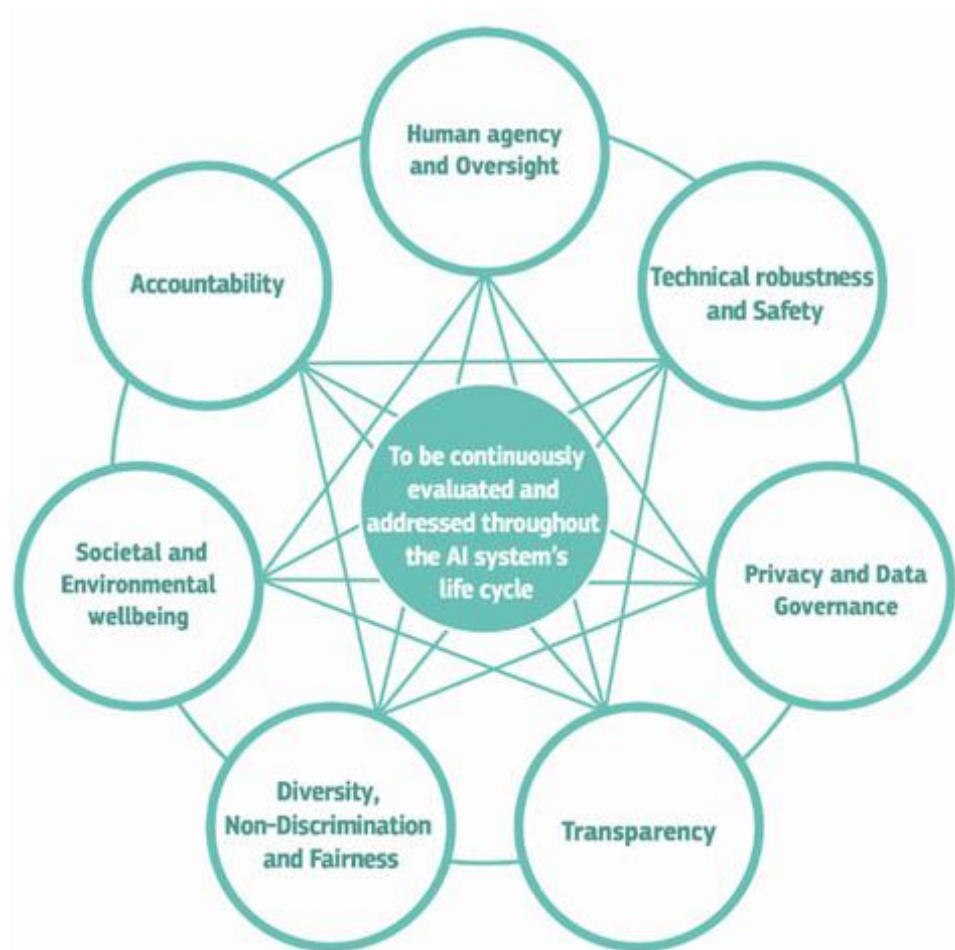
Európska únia: Akt o umelej inteligencii



Národné právne poriadky

Regulácia AI v EÚ

- Uznesenie Európskeho parlamentu zo 16. februára 2017 s odporúčaniami pre Komisiu k normám občianskeho práva v oblasti robotiky
- 2018: Komuniké umelá inteligencia pre Európu
- 2018: Koordinačný plán pre umelú inteligenciu
- AI HLEG: Etické usmernenia pre dôveryhodnú umelú inteligenciu
- 2020: Biela kniha o umelej inteligencii - európsky prístup k excelentnosti a dôvere
 - 2021: Návrh nariadenia o umelej inteligencii
 - 2022: Návrh nariadenia o zodpovednosti umelej inteligencie



Právny rámec pre AI v EÚ



- Nariadenie o AI (AIA, AI Akt)
- Nariadenie o ochrane osobných údajov (GDPR)
 - Individuálne-automatizované rozhodovanie
- Nariadenie o podpore spravodlivosti a transparentnosti pre komerčných používateľov online sprostredkovateľských služieb
 - Zverejňovanie parametrov poradia (ranking)
- Akt o digitálnych službách
 - Transparentnosť a dizajn odporúčacích systémov
- Regulácia vo finančnom sektore
- Zodpovednosť
 - Zodpovednosť za chybný výrobok
 - Zodpovednosť za škodu spôsobenú AI (stiahnuté Európskou komisiou)

Policy Area & Legislation	Industrial Policy	Connectivity	Data & Privacy	IPR	Cybersecurity	Law Enforcement	Trust & Safety	E-commerce & Consumer Protection	Competition	Media	Financial Services
European Recovery Regulation, (EU) 2021/694	Recovery and Resilience Facility Regulation, (EU) 2021/241	Frequency Bands Directive, (EEC) 1987/372	European Statistics, (EC) 2009/223, 2023/0237(COD)	Database Directive, (EC) 1996/9	Regulation for a Cybersecurity Act, (EU) 2019/881, 2023/0108(COD)	Law Enforcement Directive, (EU) 2016/680	Product Liability Directive (PLD), (EEC) 1985/374, 2022/0302(COD)	Unfair Contract Terms Directive (UCTD), (EEC) 1993/13	EC Merger regulation, (EC) 2004/139, update soon	Satellite and Cable I Directive, (EEC) 1993/83	Common European Code of Conduct, (EC) 2022/1473
European Investment Bank, (EIB) 1995/1764	InvestEU Programme Regulation, (EU) 2021/523	Radio Spectrum Decision, (EC) 2002/676	General Data Protection Regulation (GDPR), (EU) 2016/679	Community Design Directive, (EC) 2002/6, 2022/0391(COD)	Regulation to establish a European Cybersecurity Competence Centre, (EU) 2021/887	Directive on combating fraud and counterfeiting of non-cash means of payment, (EU) 2019/713	Toys Regulation, (EC) 2009/48, 2023/0290(COD)	Price Indication Directive, (EC) 1998/6	Technology Transfer Block Exemption, (EC) 2014/316	Information Society Directive, (EC) 2001/29	Administrative cooperation of tax authorities, (EU) 2021/1156
Blockchain on a pilot project, (EU) 2018/58	Connecting Europe Facility Regulation, (EU) 2021/1153	Broadband Cost Reduction Directive, (EU) 2014/61, 2023/0046(COD)	Regulation to protect personal data processed by EU institutions, bodies, offices and agencies, (EU) 2018/1725	Enforcement Directive (IPR), (EC) 2004/48	NIS 2 Directive, (EU) 2022/2555	Regulation on interoperability between EU information systems in the field of borders and visa, (EU) 2019/817	European Standardization Regulation, (EU) 2012/1025	E-commerce Directive, (EC) 2000/31	Company Law Directive, (EU) 2017/1132, 2023/0089(COD)	Audio-visual Media Services Directive (AVMSD), (EU) 2010/13	Payment Services Directive (PSD2), (EU) 2023/0001(COD)
Digital Economy & Innovation	Regulation on High Performance Computing Joint Undertaking, (EU) 2021/1173	Open Internet Access Regulation, (EU) 2015/2120	Regulation on the free flow of non-personal data, (EU) 2018/1807	Directive on the protection of trade secrets, (EU) 2016/943	Information Security Regulation, 2022/0084(COD)	Regulation on terrorist content online, (EU) 2021/784	eIDAS Regulation, (EU) 2014/910, 2021/0136(COD)	Unfair Commercial Practices Directive (UCPD), (EC) 2005/29	Market Surveillance Regulation, (EU) 2019/1020	Portability Regulation, (EU) 2017/1128	Digital Resilience Regulation, (EU) 2022/0002(COD)
	Regulation on Joint Undertakings under Horizon Europe, (EU) 2021/2085, 2022/0033(NLE)	European Electronic Communications Code Directive (EECC), (EU) 2018/1972	Open Data Directive (PSI), (EU) 2019/1024	Design Directive, 2022/0392(COD)	Cybersecurity Regulation, 2022/0085(COD)	Temporary CSAM Regulation, (EU) 2021/1232, 2022/0155(COD)	Radio Equipment Directive (RED), (EU) 2014/53	Directive on Consumer Rights (CRD), (EU) 2011/83, 2022/0147(COD)	P2B Regulation, (EU) 2019/1150	Satellite and Cable II Directive, (EU) 2019/789	Cryptocurrency Regulation, (EU) 2023/0003(COD)
	Decision on a path to the Digital Decade, (EU) 2022/2481	.eu top-level domain Regulation, (EU) 2019/517	Data Governance Act (DGA Regulation), (EU) 2022/868	Compulsory licensing of patents, 2023/0129(COD)	Cyber Resilience Act, 2022/0272(COD)	E-evidence Regulation, (EU) 2023/1543	Regulation for a Single Digital Gateway, (EU) 2018/1724	e-invoicing Directive, (EU) 2014/55	Single Market Programme, (EU) 2021/690	Copyright Directive, (EU) 2019/790	Financial Resilience Regulation, 2023/0004(COD)
	European Chips Act (Regulation), (EU) 2023/1781	Roaming Regulation, (EU) 2022/612	ePrivacy Regulation, 2017/0003(COD)	Standard essential patents, 2023/0133(COD)	Cyber Solidarity Act (Regulation), 2023/0109(COD)	Directive on combating violence against women, 2022/0066(COD)	General Product Safety Regulation, (EU) 2023/988	Geo-Blocking Regulation, (EU) 2018/302	Vertical Block Exemption Regulation (VBER), (EU) 2022/720	European Media Freedom Act, 2022/0277(COD)	Payment Services Regulation, 2023/0001(COD)
	European critical raw materials act (Regulation), 2023/0079(COD)	Regulation on the Union Secure Connectivity Programme, (EU) 2023/588	European Data Act (Regulation), 2022/0047(COD)	Digital Rights & Governance	Digital Security & Resilience	Digitalization of travel documents	Machinery Regulation, (EU) 2023/1230	Regulation on cooperation for the enforcement of consumer protection laws, (EU) 2017/2394	Digital Market Act (DMA Regulation), (EU) 2022/1925	Remuneration of musicians from third countries for recorded music played in the EU	Digital Services Act, 2023/0005(COD)
	Net Zero Industry Act, 2023/0081(COD)	New radio spectrum policy programme (RSPP 2.0)	European Health Data Space (Regulation), 2022/0140(COD)			Digital Skills & Inclusion	AI Act (Regulation), 2021/0106(COD)	Digital content Directive, (EU) 2019/770	Regulation on distortive foreign subsidies, (EU) 2022/2560	Digital Media & Culture	Regulation on combating fraud, 2023/0006(COD)
	Establishing the Strategic Technologies for Europe Platform	Digital Networks Act	Regulation on data collection for short-term				Eco-design Regulation,	Directive on certain aspects concerning contracts for the sale of	Horizontal Block Exemption Regulations (HBER)		

Legislatívny proces

- **21. Apríl 2021** – Komisia publikovala návrh nariadenia o umelej inteligencii (Artificial Intelligence Act – AIA).
- **6. December 2022** – Rada EÚ prijala pozíciu k návrhu (všeobecné smerovanie).
- **Jún 2023** – prijatie pozície od Európskeho parlamentu a začatie trialógov
- **Marec 2024** – schválenie v Európskom parlamente
- **August 2024** – nariadenie je platné



Štruktúra AI Aktu

- 113 článkov
- 180 recitálov (odôvodnení)
- 13 príloh

Dôvody prijatia

- Využívanie AI predstavuje zvýšené riziko pre bezpečnosť jednotlivcov,
- Využívanie AI predstavuje zvýšené riziko porušovania základných ľudských práv a slobôd,
- Dozorné orgány nemajú právomoci, procesné rámce a zdroje na zabezpečenie a monitorovanie súlad vývoja a používania umelej inteligencie s platnými pravidlami,
- Právna neistota ohľadom uplatnenia súčasného právneho rámca na systémy AI,
- Nedôvera k AI by znížila inovatívnosť a konkurencieschopnosť EÚ v globálnom meradle,
- Fragmentované opatrenia predstavujú prekážky pre ďalší vývoj jednotného digitálneho trhu a suverenity.

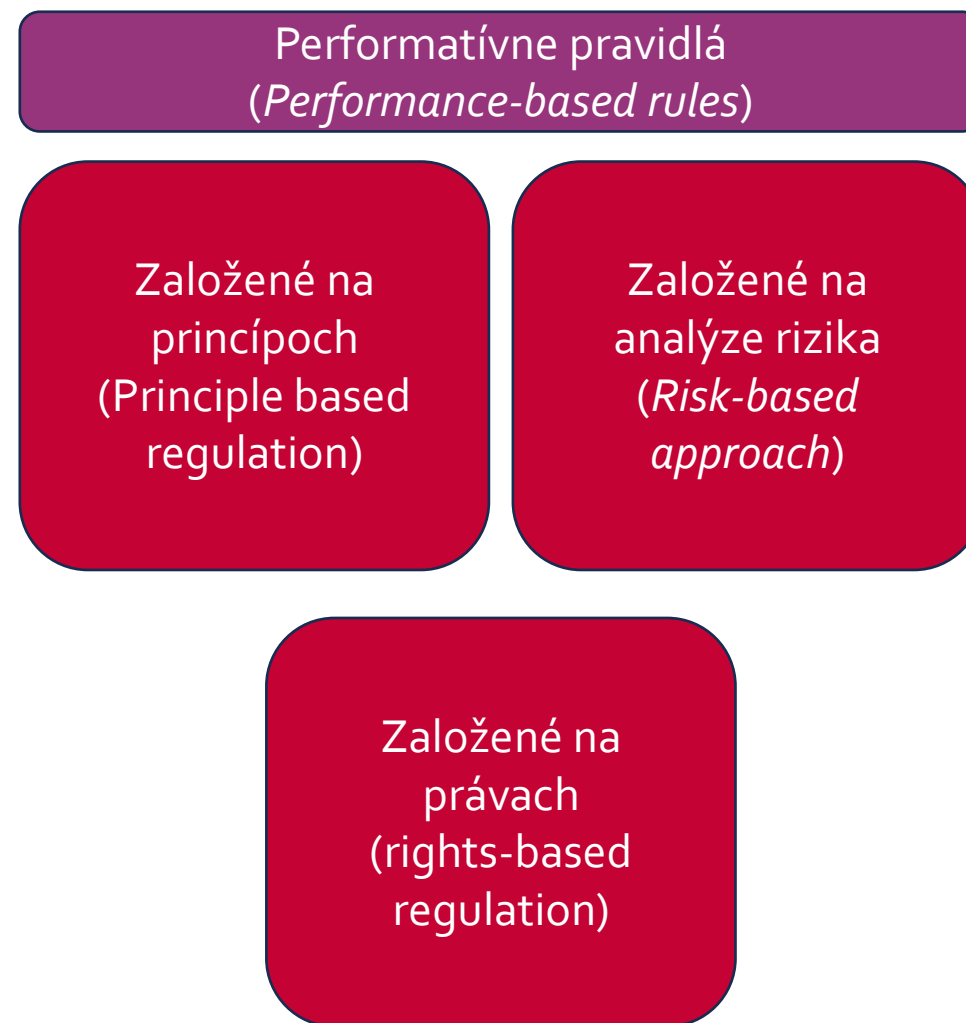
Ciele

- **Bezpečnosť**
- **Zdravie**
- **Ochrana základných práv a slobôd**
- Dôveryhodnosť AI
- Sústreďenie na človeka
- Ochrana demokracie a právneho štátu
- Udržateľnosť (článok 1)



Základné princípy

- Horizontálna regulácia
- Založená na filozofií produktového rámca
- Založená na skúmaní rizika (risk-based approach), ale upravuje aj práva pre jednotlivcov



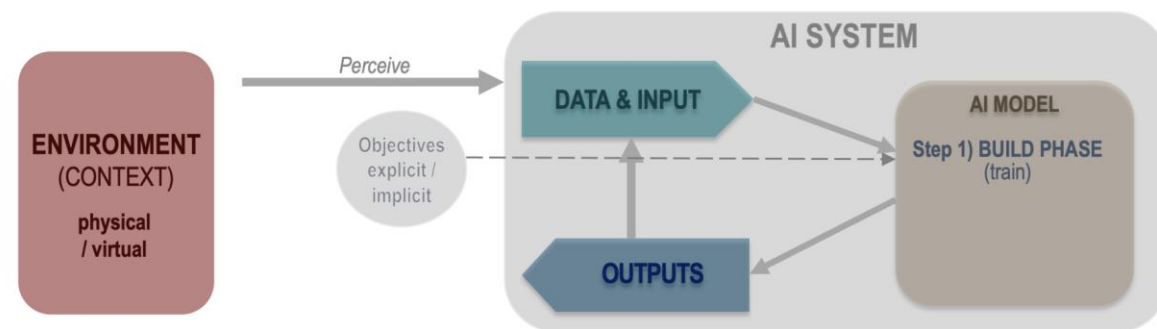
Pôsobnosť - vecná

- harmonizované pravidlá uvádzania na trh, uvádzania do prevádzky a používania systémov AI v Únii;
- zákazy určitých praktík využívajúcich AI;
- osobitné požiadavky na vysokorizikové systémy AI a povinnosti prevádzkovateľov v súvislosti s takýmito systémami;
- harmonizované pravidlá transparentnosti pre určité systémy AI;
- harmonizované pravidlá uvádzania modelov AI na všeobecné účely na trh;
- pravidlá monitorovania trhu, riadenia a presadzovania dohľadu nad trhom;
- opatrenia na podporu inovácie s osobitným dôrazom na MSP vrátane startupov.

- **Systém AI** (článok 3 bod 1)
 - je **strojový systém** dizajnovaný na prevádzku s rôznymi úrovňami **samostatnosti**, ktorý môže po nasadení prejavovať **adaptabilitu** a ktorý pre **explicitné alebo implicitné** ciele odvodzuje zo vstupov, ktoré dostáva, spôsob generovania výstupov, ako sú **predpovede, obsah, odporúčania alebo rozhodnutia**, ktoré môžu ovplyvniť fyzické alebo virtuálne prostredie;

BUILD PHASE:

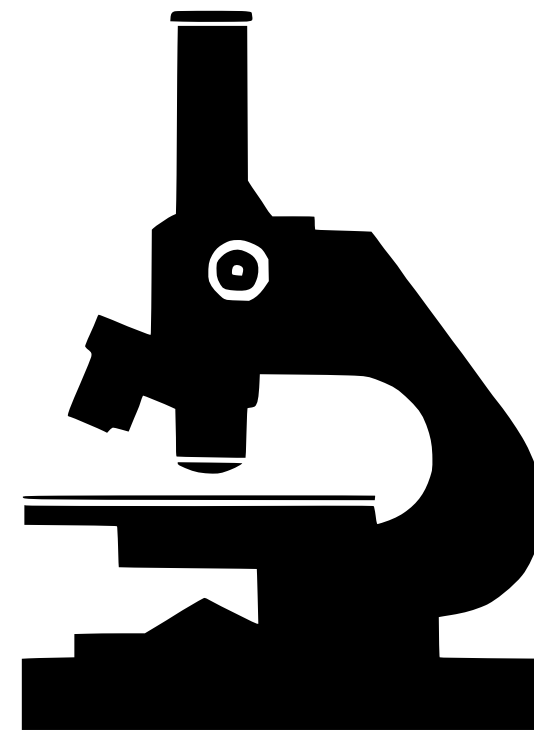
An AI system is a **machine-based** system, that



- for explicit or implicit objectives
- infers, from the input it receives
- How to generate outputs such as predictions, content, recommendations, or decisions

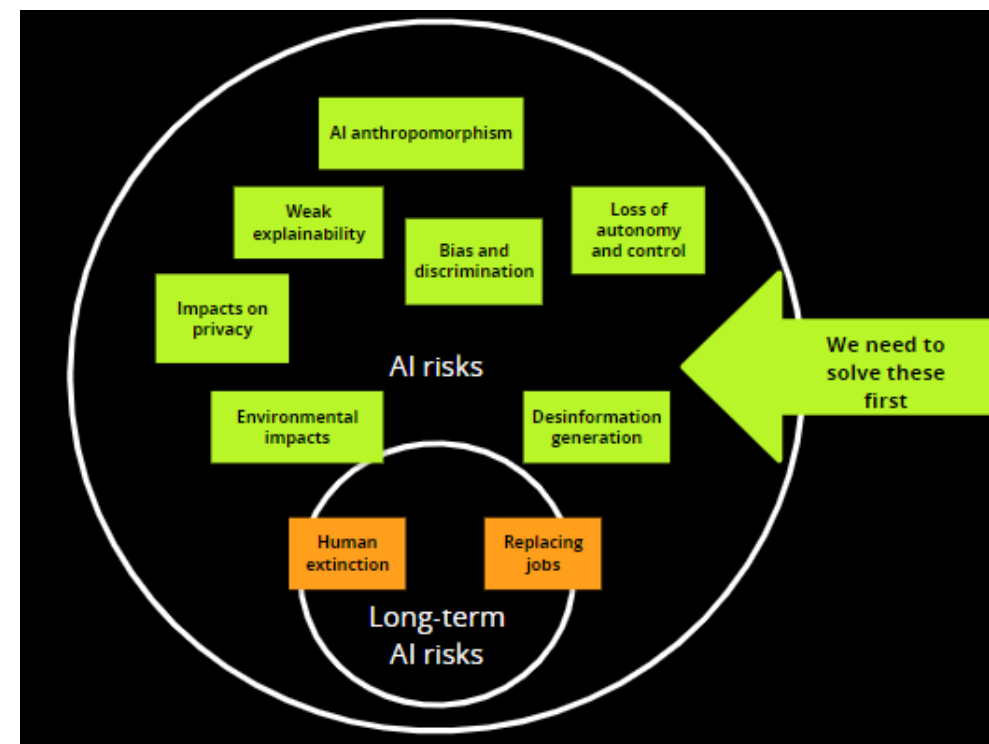
Pôsobnosť – vecná (negatívna)

- Národná bezpečnosť
- Vojenské využitie a verejná bezpečnosť
- Vedecký výstup
- Osobná činnosť (non-professional activity)



AI gramotnosť (článok 4)

- Poskytovatelia systémov AI a subjekty nasadzujúce systémy AI prijímajú opatrenia na čo najlepšie zabezpečenie dostatočnej úrovne gramotnosti v oblasti AI svojich zamestnancov a ostatných osôb, ktoré sa v ich mene zaoberajú prevádzkou a používaním systémov AI, pričom zohľadňujú ich technické znalosti, skúsenosti, vzdelanie a odbornú prípravu, ako aj kontext, v ktorom sa majú systémy AI používať, a berú do úvahy osoby alebo skupiny osôb, pri ktorých sa majú systémy AI používať.



Pôsobnosť - územná

- poskytovateľov, ktorí v Únii uvádzajú na trh alebo do prevádzky systémy AI alebo uvádzajú na trh modely AI na všeobecné účely, bez ohľadu na to, či sú títo poskytovatelia usadení alebo sa nachádzajú v Únii alebo v tretej krajine;
- subjekty nasadzujúce systémy AI, ktoré majú miesto usadenia alebo sa nachádzajú v Únii;
- poskytovateľov systémov AI a subjekty nasadzujúce systémy AI, ktoré majú miesto usadenia alebo sa nachádzajú v tretej krajine, ak sa výstup vytvorený systémom AI používa v Únii

Pôsobnosť - osobná

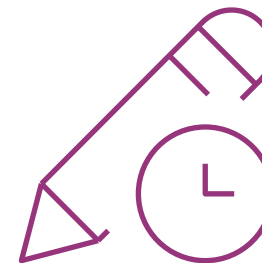
Poskytovatelia AI systémov (providers)

Systém riadenia kvality a rizík

- Správa a kvalita údajov
- Transparentnosť
- Správnosť, robustnosť, bezpečnosť
- Ľudský dohľad
- Technická dokumentácia
- Uchovávanie logov

Subjekty, ktoré AI nasadzujú (deployers)

- Posúdenie vplyvu na základné ľudské práva a slobody (iba pre niektoré subjekty)
- Súlad s požiadavkami inštrukcií od poskytovateľa
- Ľudský dohľad
- V prípade doplnenia údajov, tieto musia byť dostatočne reprezentatívne a kvalitné
- Notifikačné a monitorovacie povinnosti



Úloha

- AI systém, ktorý sa používa na monitorovanie verejných priestranstiev a identifikáciu osôb v reálnom čase na základe ich tváre. Tento systém je implementovaný mestskou samosprávou na zlepšenie bezpečnosti v mestských oblastiach, ako aj na analýzu pohybu občanov.
- AI systém vyvinutý technologickou firmou, ktorý sa používa na hodnotenie a triedenie životopisov uchádzačov o prácu na základe preddefinovaných kritérií (napr. vzdelanie, pracovné skúsenosti). Tento systém je implementovaný personálnym oddelením veľkej korporácie, ktorý si systém sama dotrénovala na historických dátach z výberových procesov.
- Systém, ktorý rozhoduje o schválení alebo zamietnutí úverových žiadostí na základe analýzy finančných údajov a kreditného skóre klienta. Tento systém bol vyvinutý špecializovanou firmou a nasadený bankou na spracovanie úverových žiadostí bez ľudského zásahu.
- Zamestnanec bez vedomia zamestnanca využíva ChatGPT na vypracovanie odpovedí na maily svojich klientov.

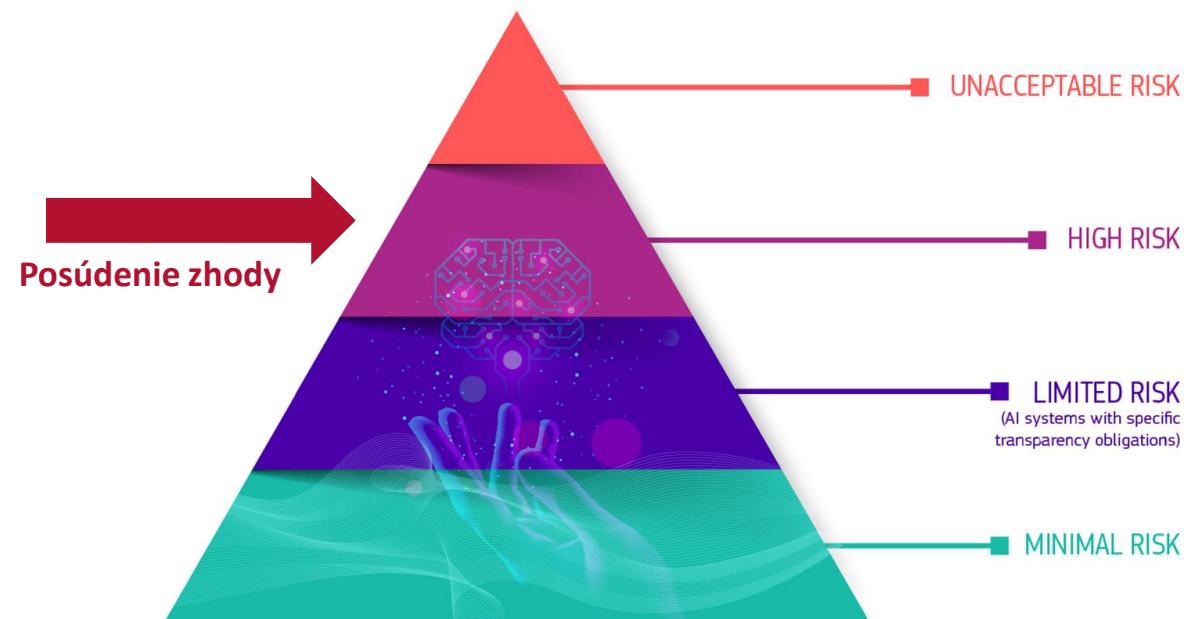
Posudzovanie na základe rizika

Neprijateľné riziko: Škodlivé AI, ktoré porušujú základné práva. Zakázané.

Vysoké riziko: Nepriaznivý vplyv na bezpečnosť ľudí alebo základné práva

Nízke riziko: Zavádzajú dobrovoľné kódexy správania a požiadavky na transparentnosť.

Minimálne riziko: Je možné dobrovoľne sa rozhodnúť uplatňovať požiadavky na dôveryhodnú AI a dobrovoľné kódexy správania.



Červené čiary

- Dve možnosti
 - Taxatívny výpočet
 - Určenie pravidiel pre posudzovanie (risk-based)



Červené čiary

Kritéria a pravidlá pre posudzovanie

- Riziko škody
- Vojenské využitie
- Potenciál pre fyzickú alebo psychologickú ujmu
- Ohrozenie základných ľudských práv a slobôd...
- Nedostatok transparentnosti a vysvetliteľnosti
- Strata kontroly

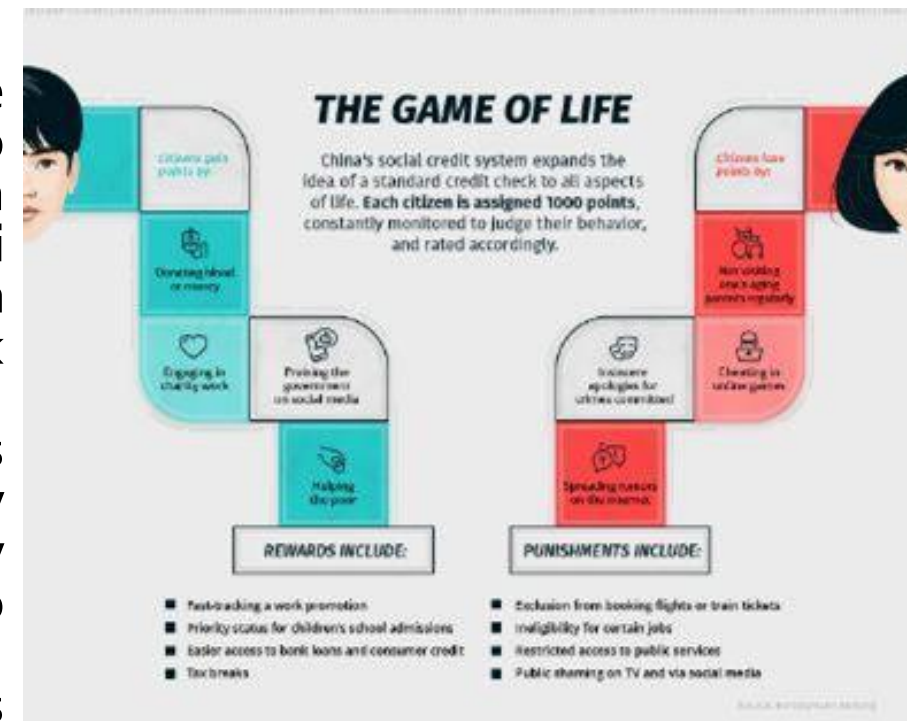
Zakázané praktiky

- Taxatívny výpočet, článok 5 AIA:
 - Biometrické kategorizačné systémy využívajúce citlivé charakteristiky
 - Sociálny skóring
 - AI systémy manipulujúce ľudské správanie
 - Odhaľovanie a zneužívanie zraniteľností
 - AI systémy na rozpoznávanie emócií na pracovisku alebo vo vzdelávaní
 - Necielený “scraping” obrázkov tvári z internetu alebo CCTV
 - Predpovedenie spáchanie trestných činov
 - Diaľková biometrická identifikácia v reálnom čase vo verejne prístupných priestoroch na účely presadzovania práva (s výnimkami)

Sociálny skóring

uvádzanie na trh, uvádzanie do prevádzky alebo používanie systémov AI na **hodnotenie alebo klasifikáciu fyzických osôb** alebo skupín osôb počas určitého obdobia na základe ich **spoločenského správania alebo známych, odvodených či predpokladaných osobných alebo osobnostných charakteristík**, pričom takto získané sociálne skóre vedie k jednému alebo obidvom z týchto výsledkov:

- i) ku škodlivému alebo nepriaznivému zaobchádzaniu s určitými fyzickými osobami alebo skupinami osôb **v sociálnych kontextoch, ktoré nesúvisia s kontextami, v ktorých boli údaje pôvodne generované alebo zhromaždené;**
- ii) ku škodlivému alebo nepriaznivému zaobchádzaniu s určitými fyzickými osobami alebo skupinami osôb, ktoré je **neodôvodnené alebo neprimerané ich spoločenskému správaniu alebo jeho závažnosti;**



Manipulovanie ľudského správania

- uvádzanie na trh, uvádzanie do prevádzky alebo používanie systému AI, ktorý využíva podprahové techniky mimo vedomia osoby alebo účelovo manipulatívne alebo klamlivé techniky s cieľom alebo účinkom podstatne narušiť správanie osoby alebo skupiny osôb tým, že sa citeľne oslabí ich schopnosť prijať informované rozhodnutie, čo zapríčini, že prijmú rozhodnutie, ktoré by inak neprijali, a to spôsobom, ktorý tejto osobe, inej osobe alebo skupine osôb spôsobuje alebo pri ktorom je odôvodnené predpokladať, že pravdepodobne spôsobí značnú ujmu;

Úloha

- Systém využíva algoritmy strojového učenia na sledovanie produktivity zamestnancov v reálnom čase a na hodnotenie ich výkonu na základe rôznych metrík, ako sú čas strávený pri úlohách, rýchlosť vykonávania úloh a reakčné časy na požiadavky nadriadených.
- AI systém analyzuje historické údaje o klientoch a predikuje ich pravdepodobnosť splatenia úveru. Na základe týchto analýz systém automaticky rozhoduje o schválení alebo zamietnutí úverov, pričom je využívaný bez ľudského zásahu.
- AI systém automaticky monitoruje príspevky na sociálnych médiách a označuje obsah ako dezinformácie alebo nelegálny, bez prítomnosti ľudského overenia. Tento systém je schopný automaticky odstrániť príspevky, ktoré považuje za nevyhovujúce.
- AI systém sleduje správanie občanov v mestských oblastiach prostredníctvom kamier a senzorov. Na základe správania (napr. prejavy agresie, podozrivé správanie) systém upozorňuje bezpečnostné zložky, ktoré môžu zasiahnuť.

Ako zistím, či je systém AI vysokorizikový?

- Systém AI je určený na používanie ako bezpečnostný komponent výrobku (alebo je sám takýmto výrobkom), na ktorý sa vzťahujú harmonizačné predpisy EÚ (**Príloha I**)
- výrobok, ktorého bezpečnostným komponentom je systém AI, alebo samotný systém AI ako výrobok sa musí podrobiť posúdeniu zhody treťou stranou (**Príloha I**)
- Systémy AI uvedené v **Prílohe III**
- **systém AI nepovažuje za vysokorizikový, ak nepredstavuje významné riziko ujmy na zdraví, bezpečnosti alebo základných právach fyzických osôb, a to ani tým, že podstatným spôsobom neovplyvňuje výsledok rozhodovania**

**Oblasť, v ktorej sa bude používať
vysokorizikový systém AI****Biometrická
identifikácia****Kritická
infraštruktúra****Vzdelávanie a
príprava****Zamestnanosť,
riadenie pracovníkov a
prístup k samostatnej
zárobkovej činnosti****Prístup k základným
súkromným službám a
základným verejným
službám****Migrácia, azyl a
riadenie kontroly
hraníc****Presadzovanie
práva****Výkon spravodlivosti a
demokratické procesy****Odkaz na osobitné nariadenie****Nariadenie o zdravotníckych pomôckach****Nariadenie o strojníckych výrobkoch****Nariadenie o bezpečnosti hračiek****a ďalšie...**

RETAIL

Amazon scraps a secret A.I. recruiting tool that showed bias against women

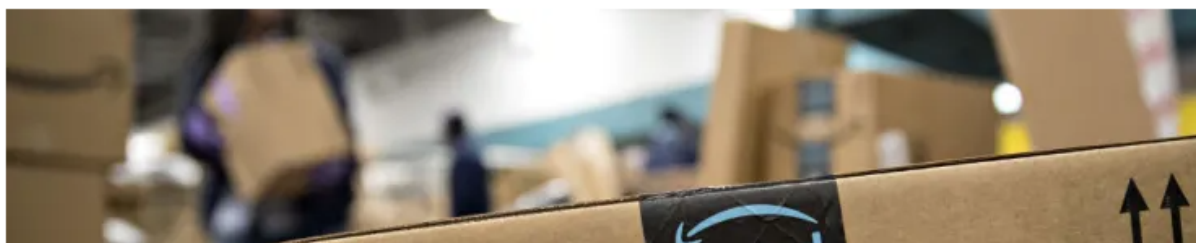
PUBLISHED WED, OCT 10 2018 6:15 AM EDT | UPDATED THU, OCT 11 2018 2:25 PM EDT

REUTERS

SHARE [f](#) [X](#) [in](#) [✉](#)

KEY POINTS

- Amazon.com's machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.
- The team had been building computer programs since 2014 to review job applicants' resumes with the aim of mechanizing the search for top talent, five people familiar with the effort told Reuters.
- The company's experimental hiring tool used artificial intelligence to give job candidates scores ranging from one to five stars — much like shoppers rate products on Amazon, some of the people said.



CNBC TV

**Fast Money
Halftime Report**

WATCH LIVE

UP NEXT | **The Exchange** 01:00 pm ET

[Listen](#)

TRENDING NOW



Stocks making the biggest moves midday: Nvidia, Super Micro Computer, Unilever, Nordstrom and more



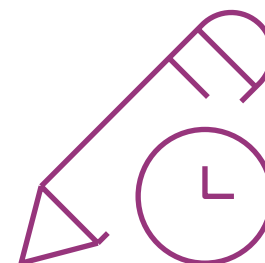
Nvidia stock dips after company unveils latest AI chips



CLUB We're looking to buy more of this health care stock after an

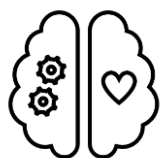
Úloha

- Zohľadnite kontext kauzy Amazonu a odpovedzte na nasledujúce otázky
 - Je možné daný AI systém považovať za vysokorizikový?
 - Aké požiadavky v zmysle AI Aktu problémy identifikované vo videu?



Posúdenie zhody (conformity assessment)

KROK 1



Vysokorizikový
system AI je vo
vývoji

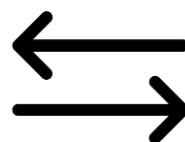
KROK 2



Vykonanie
**posúdenia
zhody**



KROK 3



Registrácia v
databáze
vysokorizikovýc
h systémov AI

KROK 4



Deklarácia o
zhode a
označenie CE

**Systém
môže byť
uvedený na
EÚ trhu**

KROK 5



Monitorovanie
na trhu

Povinnosti pre poskytovateľov (článok 9)

Požiadavky na vysokorizikové systémy AI (článok 16)

Potrebná
dokumentácia

...

Vyhlásenie o
zhode

Systém riadenia kvality (článok 17)

Systém
riadenia rizík

Stratégia dodržiavania
regulačných
požiadaviek

Monitorovanie
na trhu

...

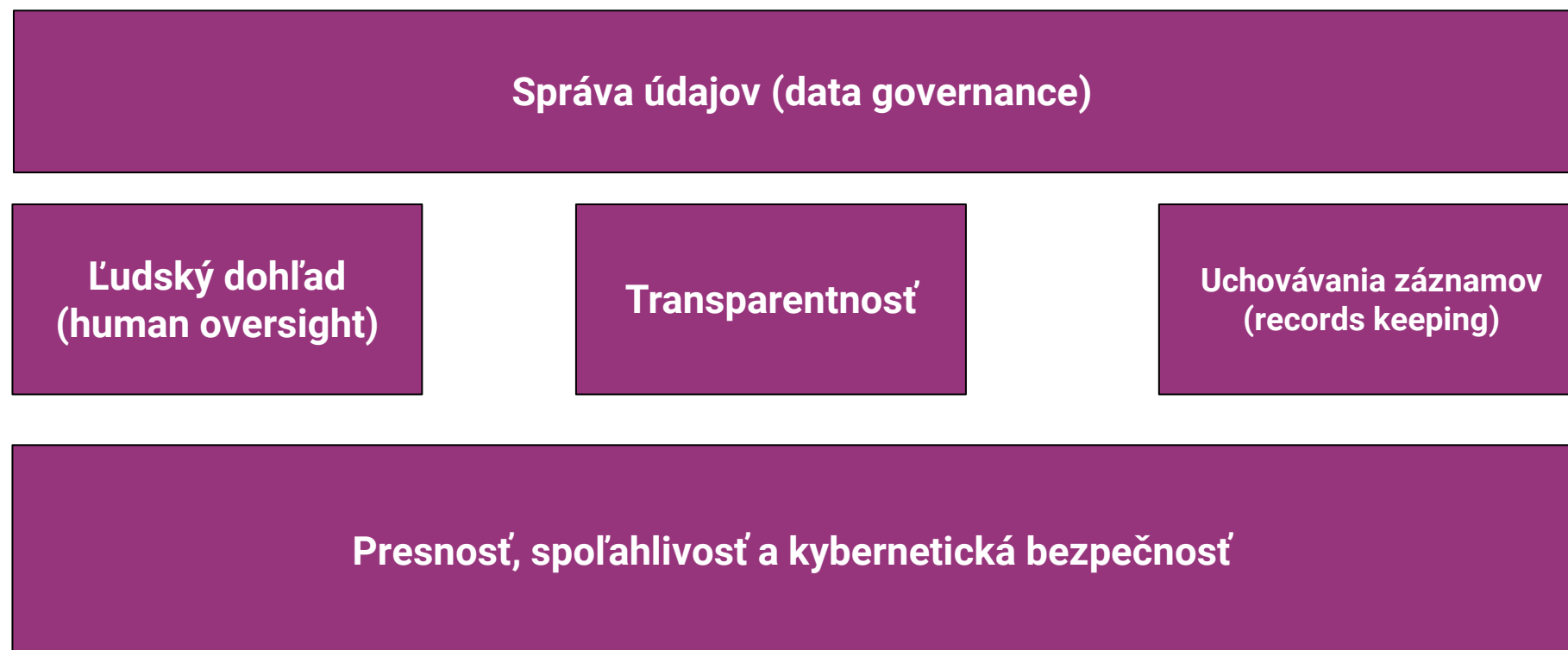
System riadenia rizík

- Proces
 - Identifikácia a analýza známych a odôvodnene predvídateľných rizík
 - Odhad a hodnotenie rizík
 - Hodnotenie rizík získaných pri monitorovaní systému AI na trhu
 - Prijatie vhodných a cielených opatrení na riešenie rizík

Súlاد s AIA

- Prostredníctvom posúdenia zhody
 - postup preukázania, či boli splnené požiadavky stanovené v kapitole II oddiele 2 týkajúce sa vysokorizikového systému AI
 - Vnútorňá kontrola
 - S pomocou notifikačného orgánu
- Prostredníctvom harmonizovaných štandardov

Požiadavky na vysokorizikové systémy AI



Subjekty, ktoré AI nasadzujú (deployers)

- Posúdenie vplyvu na základné ľudské práva a slobody (iba pre niektoré subjekty)
- Súlad s požiadavkami inštrukcií od poskytovateľa
- Ľudský dohľad
- V prípade doplnenia údajov, tieto musia byť dostatočne reprezentatívne a kvalitné
- Notifikačné a monitorovacie povinnosti

Posúdenie vplyvu na základné ľudské práva a slobody

Plánovanie a rozsah

- Opis produktu a systému
- Kontext ľudských práv
- Identifikácia držiteľov práv a zainteresovaných osôb



Zber údajov a analýza rizík

- Identifikácia konkrétnych ľudských práv
- Posúdenie pravdepodobnosti
- Posúdenie závažnosti zásahu
- Posúdenie celkového vplyvu
- Opatrenia na zmiernenie

Table 2.2 Probability

	Probability	
Low	The risk of prejudice is improbable or highly improbable	1
Medium	The risk may occur	2
High	There is a high probability that the risk occurs	3
Very high	The risk is highly likely to occur	4

Source The author

Table 2.3 Exposure

	Exposure	
Low	Few or very few of the identified population of rights-holders are potentially affected	1
Medium	Some of the identified population are potentially affected	2
High	The majority of the identified population is potentially affected	3
Very high	Almost the entire identified population is potentially affected	4

Source The author

Table 2.4 Likelihood table (L)

		Probability				Likelihood	
		1	2	3	4		
Exposure	1	1	2	3	4	Low	1
	2	2	3	5	9	Medium	2
	3	3	5	9	12	High	3
	4	4	7	12	15	Very high	4

Source The author

Table 2.5 Gravity of the prejudice

	Gravity of the prejudice	
Low	Affected individuals and groups may encounter only minor prejudices in the exercise of their rights and freedoms	1
Medium	Affected individuals and groups may encounter significant prejudices	2
High	Affected individuals and groups may encounter serious prejudices	3
Very high	Affected individuals and groups may encounter serious or even irreversible prejudices	4

Source The author

Table 2.6 Effort to overcome the prejudice and to reverse adverse effects

	Effort	
Low	Suffered prejudice can be overcome without any problem (e.g. time spent amending information, annoyances, irritations, etc.)	1
Medium	Suffered prejudice can be overcome despite a few difficulties (e.g. extra costs, fear, lack of understanding, stress, minor physical ailments, etc.)	2
High	Suffered prejudice can be overcome albeit with serious difficulties (e.g. economic loss, property damage, worsening of health, etc.)	3
Very high	Suffered prejudice may not be overcome (e.g. long-term psychological or physical ailments, death, etc.)	4

Source The author

Table 2.7 Severity table (S)

		Gravity				Severity	
		1	2	3	4		
Effort	1	1	2	4	6	Low	1
	2	2	3	5	8	Medium	2
	3	3	5	8	10	High	3
	4	5	8	10	12	Very high	4

Source The author

Table 2.11 Table of envisaged risks for the examined case (L: low, M: medium; H: high; VH: very high)

Risk	L	S	Overall impact
Impact on privacy and data protection	VH	M	H
Impact on freedom of thought	M	L	M
Impact on the right to psychological and physical safety	L	M	M

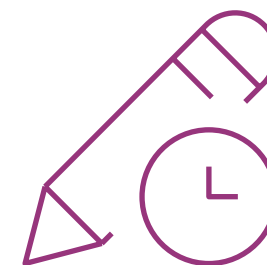
Source The author

Relevantné práva a slobody pre AI

- Ľudská dôstojnosť
- Právo na súkromie
- Právo na ochranu osobných údajov
- Zákaz diskriminácie a rovnosť
- Právo na spravodlivý súdny proces
- Právo na dobrú správu vecí verejných
- Sloboda prejavu
- Právo na priaznivé životné prostredie
- Sloboda zhromažďovania

Relevantné práva a slobody pre AI

- Aké práva budú relevantné pri nasadení AI v sektore zdravotníctva pri diagnostike a predikcii rakovinotvorných buniek?
 - Vyberte si jedno z práv a skúste charakterizovať
 - Pravdepodobnosť (probability) vplyvu naň
 - Kvantitu vystavenia (exposure)
 - Závažnosť vplyvu na jednotlivcov (gravity of prejudice)
 - Mieru, ktorú musia jednotlivci vynaložiť na prekonanie vplyvu (effort to overcome)



Modely na všeobecné účely

- model AI, a to aj trénovaný veľkým množstvom údajov s použitím samokontroly vo veľkom rozsahu, ktorý má významnú všeobecnú povahu a je schopný kompetentne vykonávať širokú škálu odlišných úloh bez ohľadu na spôsob, akým sa model uvádza na trh, a ktorý možno integrovať do rôznych nadväzujúcich systémov alebo aplikácií, s výnimkou modelov AI, ktoré sa využívajú na účely výskumných, vývojových a prototypových činností pred ich vydaním na trh (článok 3 bod 63)



Modely na všeobecné účely

MODEL AI NA VŠEOBECNÉ POUŽITIE

- Požiadavky transparentnosti:
 - Koncipovanie technickej dokumentácie
 - Súlad s autorskými právami a prehľad využitých zdrojov na tréovanie

MODEL AI NA VŠEOBECNÉ POUŽITIE SO SYSTÉMICKÝM RIZIKOM

- Požiadavky na modely a systémy AI na všeobecné použitie
- Hodnotenie modelu z pohľadu state-of-the-art
- Analýza a zmierňovanie rizík
- Reportovanie incidentov
- Primeraná úroveň kybernetickej bezpečnosti

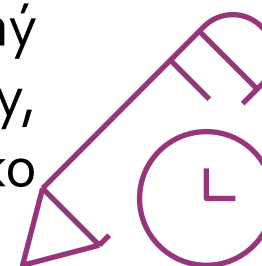
Monitorovanie po uvedení na trh

- aktívne a systematicky zbierajú, dokumentujú a analyzujú relevantné údaje o výkonnosti vysokorizikových systémov AI počas celej ich životnosti, ktoré môžu poskytovať nasadzujúce subjekty alebo ktoré možno zbierať prostredníctvom iných zdrojov a ktoré umožňujú poskytovateľovi hodnotiť nepretržitý súlad systémov AI s požiadavkami stanovenými AI Akte
- Podávanie správ o závažných incidentoch



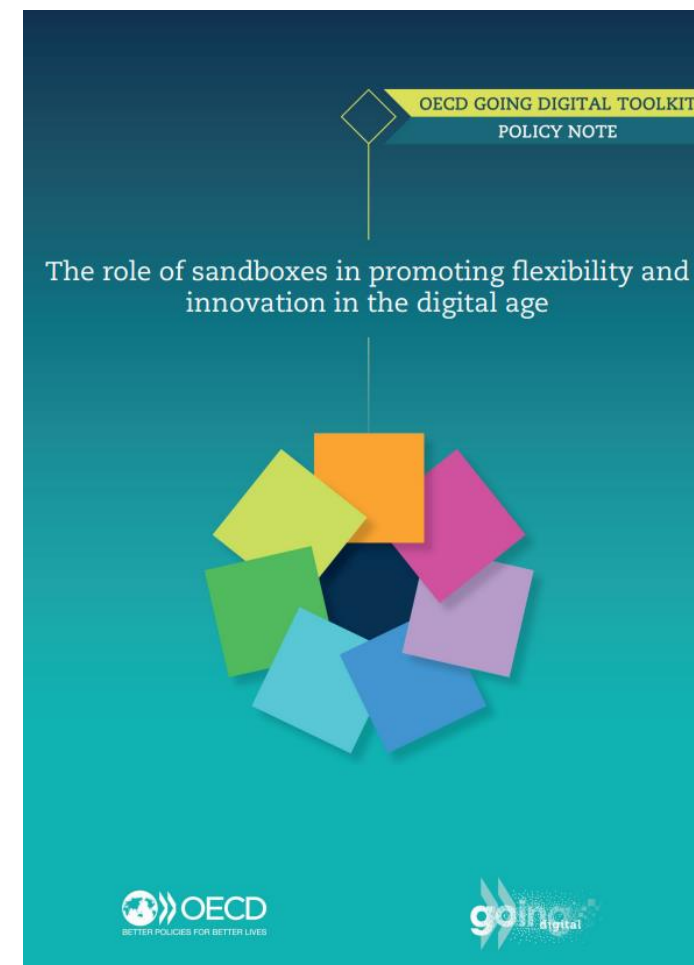
Transparentnosť

- Priama interakcia AI systému s užívateľom
- Rozpoznávanie emócií alebo systému biometrickej kategorizácie
- Umelo generovaný obsah
- Transparentné
- Strojovo čitateľné (watermarks)
- Transparentnosť deepfakes
 - deep fake je obrazový obsah, audioobsah alebo videoobsah vytvorený alebo zmanipulovaný pomocou AI, ktorý sa podobá na existujúce osoby, objekty, miesta či iné subjekty alebo udalosti a osobe by sa falošne javil ako autentický alebo pravdivý



Podpora inovácii – experimentálne regulované prostredie

- Aspoň jedno za každý členský štát
- Kontrolované prostredie, ktoré podporuje inováciu a uľahčuje vývoj, tréning, testovanie a validáciu inovačných systémov AI na obmedzený čas pred ich uvedením na trh alebo do prevádzky podľa konkrétneho plánu daného experimentálneho prostredia dohodnutého medzi potenciálnymi poskytovateľmi a príslušným orgánom
- Usmernenia dozorných orgánov
- Možnosť aplikovať AI systém aj v reálnych podmienkach



Práva osôb

- Právo podať sťažnosť (článok 85)
- Právo na vysvetlenie individuálneho rozhodovania (článok 86)

Dohľad



Výbor pre AI

Úrad pre AI

Poradné fórum

Vedecký panel



Dozorný orgán

Účinnosť

- **Zakázané postupy** majú obdobie na dosiahnutie súladu 6 mesiacov.
- **AI modely na všeobecné účely** majú 12 mesiacov; 24 mesiacov, ak sú už na trhu
- **Vysoko rizikové systémy** podľa prílohy III majú 24 mesiacov; systémy, ktoré sú v tejto chvíli na trhu sa vôbec nevzťahuje, pokiaľ sa na nich nevykonajú **významné zmeny**
- **Vysoko rizikové systémy** uvedené v prílohe I majú 36 mesiacov; systémy, ktoré sú v tejto chvíli na trhu sa vôbec nevzťahuje, pokiaľ sa na nich nevykonajú **významné zmeny**

PRÁVNE NÁSTROJE BOJA PROTI DEZINFORMÁCIÁM V ONLINE PRIESTORE

Dezinformácie a kybernetická bezpečnosť

- Dezinformácie môžu byť nástrojom kybernetických útokov a hrozieb, ktoré ovplyvňujú politickú stabilitu a verejnú dôveru.
- Kybernetické útoky často využívajú dezinformácie na šírenie strachu, zmätku a polarizácie medzi verejnosťou.
- Manipulácia s informáciami v online priestore môže byť súčasťou širších hybridných hrozieb, ktoré kombinujú tradičné a digitálne spôsoby útokov.
- Ovládanie verejnej mienky prostredníctvom dezinformácií môže viesť k destabilizácii spoločenských a politických štruktúr.

Hybridné hrozby a dezinformácie

- Hybridné hrozby kombinujú rôzne formy agresie: kybernetické, informačné, politické a vojenské útoky.
- Dezinformácie sú kľúčovým nástrojom v hybridných hrozbách, keďže môžu ovplyvniť verejné názory a vyvolať sociálne napätie.
- Cielené šírenie nepravdivých informácií môže slúžiť ako prostriedok na manipuláciu s verejnosťou, čo zvyšuje zraniteľnosť krajiny voči vonkajším alebo vnútorným hrozbám.

Politiky na úrovni Európskej únie

- **Akčný plán EÚ proti dezinformáciám (2018):** Sústredený na zlepšenie koordinácie medzi členskými štátmi a kľúčovými aktérmi, ako sú sociálne médiá a technologické firmy.
- **Kódex správania proti dezinformáciám:** Dobrovoľná iniciatíva veľkých online platforiem, ktorá zaručuje transparentnosť a zodpovednosť za šírenie dezinformácií.

Akčný plán boja proti dezinformáciám (2018)

- Cieľom je posilniť spôsobilosť a spoluprácu EÚ v boji proti dezinformáciám
- **Kľúčové opatrenia:**
 - Podpora dobrovoľných iniciatív ako Kódex správania proti dezinformáciám.
 - Posilnenie spolupráce s online platformami na transparentnosti a odhaľovaní falošných informácií.
 - Podpora médií a nezávislých novinárov, ktorí bojujú proti dezinformáciám.
 - Zavedenie rýchlej reakcie na krízové situácie, ako sú volebné obdobia a obdobia zraniteľnosti.

Akčný plán pre európsku demokraciu (2020)

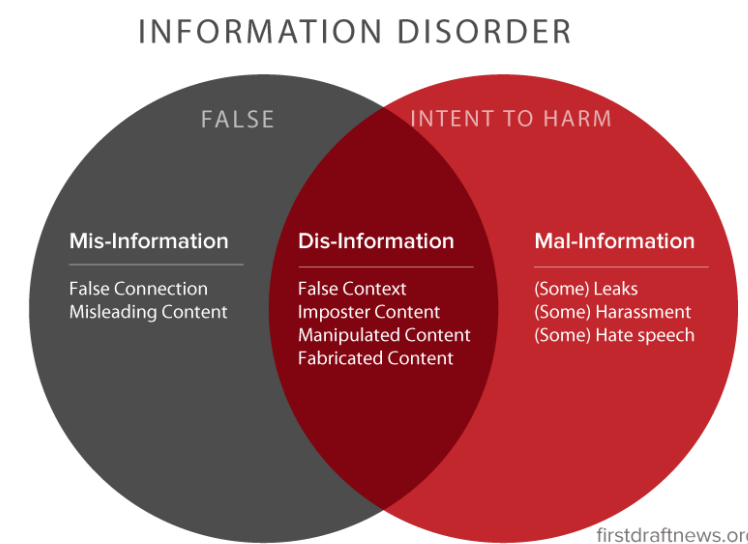
- Cieľom je chrániť demokratické procesy pred manipuláciami, vrátane dezinformácií a hybridných hrozieb.
- **Kľúčové opatrenia:**
 - Zabezpečenie väčšej transparentnosti online politických reklám.
 - Podpora politickej plurality a ochrany slobodných volieb pred vplyvom cudzích aktérov.
 - Zavedenie opatrení na ochranu nezávislosti médií a novinárov.
 - Posilnenie právnej a inštitucionálnej základne pre boj proti vonkajším zasahovaním do demokratických procesov.

Slovensko

- **Koncepcia strategickej komunikácie SR (2023)**
 - Schválená vládou odborníkov, zameraná na zlepšenie informovanosti občanov o členstve v EÚ a NATO.
 - Cieľom je zvýšiť dôveru v štát a jeho demokratické inštitúcie, čím sa zamedzí šíreniu dezinformácií a konšpiračných teórií.
- **Koncepcia Slovenskej republiky pre boj proti hybridným hrozbám (2018)**
 - Dokument vypracovaný Národným bezpečnostným úradom (NBU), ktorý definuje hybridné hrozby ako súbor nátlakových a podvratných činností, konvenčných a nekonvenčných, vojenských a nevojenských metód, ktoré môžu štátne aj neštátne subjekty koordinovaným spôsobom využívať na dosiahnutie konkrétnych cieľov bez formálneho vyhlásenia vojny a pod prahom zvyčajnej reakcie.

Čo je to dezinformácia?

- Nie je definovaný v právnom poriadku Slovenskej republiky
 - Akčný plán pre európsku demokraciu: „Dezinformácie sú **nepravdivý alebo zavádzajúci obsah**, ktorý sa šíri s **úmyslom** klamať alebo zabezpečiť hospodársky alebo politický zisk a ktorý môže spôsobiť ujmu verejnosti.“
- Množstvo podobných pojmov
 - Misinformácia
 - Malinformácia
 - Falošná správa
 - Hoax



Sloboda prejavu a právo na informácie

- Priama regulácia dezinformácií
- Regulovanie údajov
- Regulovanie prostredia



Faktory ovplyvňujúce šírenie dezinformácií

Psychologické faktory

Kognitívne faktory

- intuitívne premýšľanie, kognitívne zlyhania a iluzórna pravda

Sociálne afektívne faktory

- zdroj informácie, emócie a názory a viera adresáta informácie

Sociologické faktory

- Úbytok sociálneho kapitálu
- Nárast nerovnosti
- Zvyšujúcu sa polarizáciu spoločnosti
- Nedôvera vo vedu
- Evolúciu mediálneho prostredia

Ekonomické faktory

- Ekonomika pozornosti
- Dohľad kapitalizmu

Technické faktory

- Personalizácia obsahu odporúčacími systémami (AI)
- Filtračné bubliny

Nástroje verejného práva pre boj s dezinformáciami online



Akt o digitálnych službách (DSA)

Akt o umelej inteligencii (AIA)

Nariadenie o ochrane osobných údajov (GDPR)



Zákon o mediálnych službách

Zákon o kybernetickej bezpečnosti

Priestupkový zákon

Tajné služby

Trestný zákon

Návrh zákona o zodpovednosti a transparentnosti online platforiem

Akt o digitálnych službách



- **Nezákonný** obsah
- Dezinformácie v pôsobnosti DSA
 - Ak sa tak **rozhodne** platforma
 - Ak sú dezinformácie **nezákonné**
- Okrem vyššie uvedených prípadov sú aj predmetom **posudzovania rizík a auditovania** veľmi veľkých online platforiem

Požiadavky na veľmi veľké online platformy

- Transparentnosť a dizajn odporúčacích systémov
- Reportovacie povinnosti
- Analáza a zmierňovanie rizík
- Auditovanie
- Ochrana maloletých
- Dizajn platforiem
- Pravidlá pre zobrazovanie reklamy
- Prístup pre výskumníkov
- Dôveryhodní nahlasovatelia

Akt o umelej inteligencii



- Je to **zakázaná** praktika?
- Sú odporúčacie systémy klasifikované ako **systémy AI vysokého rizika**?
- Pravidlá pre **modely AI na všeobecné účely**
- Požiadavky na **transparentnosť**

Požiadavky na systémy AI vysokého rizika

- Správa údajov
- Dokumentácia k systému
- Transparentnosť
- Ukladanie záznamov
- Ľudský dohľad
- Odolnosť, presnosť, reprezentatívnosť
- Kybernetická bezpečnosť

Všeobecné nariadenie o ochrane osobných údajov



- Relevancia pre dezinformácie
 - Odporúčacie systémy využívajú **profilovanie**
 - **Práva** dotknutých osôb
- Požiadavka na **špecificky navrhnutú ochranu osobných údajov a vykonanie posúdenia vplyvu**
- Pravidlá pre **automatizované individuálne rozhodovanie**

Práva dotknutých osôb

- Právo na informácie
- Právo na prístup
- Právo na opravu
- Právo na vymazanie
- Právo na prenosnosť
- Právo namietat'
- Právo nebyť predmetom automatizovaného individuálneho rozhodovania

Zákon o mediálnych službách



- Osobitný typ správneho konania
 - Konanie o **zamedzení šírenia nelegálneho obsahu**
- **Úzko vymedzený nelegálny obsah:** detská pornografia, teroristický obsah, podnecovanie násilia alebo nenávistné prejavy a extrémistické materiály

Rozhodnutia o začatí konania

- 53 podnetov
- 2 prípady, keď konanie začalo
- 8 prípadov keď bol obsah vymazaný pred rozhodnutím o začatí konania

Zákon kybernetickej bezpečnosti



- Blokovanie webstránok
 - Škodlivá aktivita vrátane závažných dezinformácií
- Zablokované 4 webstránky

Problematické aspekty

- Nebola dodržaná rovnosť zbraní
- Nedostatočný nezávislý dohľad
- Neefektívne opravné prostriedky
- Nedostatočná transparentnosť
- Notifikácia vopred

Priestupkový zákon

- Falšovanie informácií o zdravotnom stave (§20/1/c)
- Vzbudenie **verejného pohoršenia** (§47/1/c)
- **Ublíženie na cti** (§49/1/a)
- Vyhrážanie (§ 49/1/d)

Potenciálne problémy

- Nedbanlivosť neseď do kontextu šírenia dezinformácií
- Niektoré priestupky sa viažu na veľmi špecifický kontext
- Niektoré priestupky si budú vyžadovať extenzívnu interpretáciu

Trestný zákon



- Ohováranie
- Šírenie poplašnej správy
- Ohrozenie bezpečnosti vzdušného dopravného prostriedku a lode
- Falšovanie a vyhotovenie nepravdivej zdravotnej dokumentácie
- Teroristický útok

Potenciálne problémy

- Prostriedok *ultima ratio*
- *Návrh trestného činu šírenia nepravdivých správ*

Právomoci tajných služieb

- Slovenská informačná služba
 - Dezinformácie nie sú explicitne upravené v zákone, avšak **môžu byť predmetom vyšetrovania a prijímania opatrení**
- Vojenské spravodajstvo
 - Vyhodnocovanie informácií v kontexte **hybridných hrozieb a dezinformácií**



**NO IMAGE
AVAILABLE**

Vyhodnocovanie proporcionality

- Metriky
 - Koordinovaný mechanizmus odolnosti Slovenskej republiky voči informačným operáciám
 - disinfohub.eu
- Dané do kontextu rozhodnutí slovenských súdov



- Angažovanosť
- Vystavenie
- Počet platforiem
- Počet jazykov
- Dosah na mainstreamové media
- Typ aktéra
- Formát
- Výzva na konanie
- Cieľ dezinformácie
- Vec verejného záujmu

1. Engagement on social networks: measuring shares and reactions

This measure takes into consideration the social media platform where the single piece of disinformation collected the highest number of **shares and reactions**, regardless of its appearance in more than one platform (which will be considered later). In fact, reactions and shares imply that there has been user's active **engagement** with the content.

- 0 – 1.000 shares and reactions = 0 points
- 1.001 - 10.000 shares and reactions = 1 point
- 10.001 – 100.000 shares and reactions = 2 points
- More than 100.001 shares and reactions = 3 points

2. Exposure on social networks: measuring the views

This category measures the **exposure** that users had to the content in the form of views, although we acknowledge the impossibility to fully assess whether they intentionally decided to play a video, or it automatically played. Therefore, this category, which is to be counted only for the platform gathering the highest metrics, completes the previous one.

- 0 – 1.000 views = 0 points
- 1.001 - 10.000 views = 1 point
- More than 10.001 views = 2 points

3. Content circulation: the number of platforms

This indicator is intended to measure cross-platform diffusion, giving one extra point if the content has been shared on more than two platforms, including messaging chat services such as Telegram (due to the social network similar function of the channels) and WhatsApp (whose role in spreading disinformation cannot be ignored).

- Content shared on one or two platforms = 0 points
- Content shared on more than two platforms = 1 point

- Na základe vyhodnotenia by bolo možné dezinformáciu alebo kampaň klasifikovať podľa vplyvu na:
 - Alarmujúci
 - Vysoký
 - Stredný
 - Nízky
- Ktorý nástroj je vhodný pri ktorom vplyve?

