

**Kempelenov
inštitút
inteligentných
technológií**



**Analýza vybraných právnych
úprav navrhovaných Európskou
komisiou a technologických riešení
vo vzťahu k šíreniu dezinformácií a
správaniu sa online platforiem**

Analýza vybraných právnych úprav navrhovaných Európskou komisiou a technologických riešení vo vzťahu k šíreniu dezinformácií a správaniu sa online platforiem

Táto štúdia bola vypracovaná Kempelenovým inštitútom inteligentných technológií na základe zadania poslankyne Európskeho parlamentu Miriam Lexmann. Zmluva o poskytovaní služieb bola uzavretá dňa 1.10.2021.

Autormi štúdie sú Matúš Mesarčík, Róbert Móro, Michal Kompan, Juraj Podroužek, Jakub Šimko a Mária Bieliková z Kempelenovho inštitútu inteligentných technológií.

Vzor citácie:

Mesarčík, M., Móro, R., Kompan, M., Podroužek, J., Šimko, J., Bieliková, M. Analýza vybraných právnych úprav navrhovaných Európskou komisiou a technologických riešení vo vzťahu k šíreniu dezinformácií a správaniu sa online platforiem. Kempelenov inštitút inteligentných technológií. Marec 2022.

Manažérske zhrnutie

Slobodný prístup k informáciám, sloboda slova a schopnosť vyhodnocovať pravdivosť informácií a prijímať na základe nich rozhodnutia patria k základným pilierom demokratických spoločností. Nové technológie vrátane sociálnych médií a iných online platforiem dali každému z nás nebyvalé možnosti vyjadrovať svoje názory a oslovovať nimi potenciálne veľké publikum. Na druhej strane, s príspevím systémov umelej inteligencie, predovšetkým personalizovaných odporúčacích systémov, online platformy napomáhajú šíriť aj nepravdivé informácie vrátane dezinformácií s významnými negatívnymi dopadmi na spoločnosť.

Vzhľadom na tieto negatívne dopady na jednej strane a na silný rešpekt k základným ľudským právam a slobodám zakotvený v európskych právnych systémoch na strane druhej, predstavujú dezinformácie jednu z najväčších výziev pre súčasné iniciatívy pri tvorbe pravidiel, ktorých cieľom je regulovať obsah a zodpovednosť na internete.

Ambíciou predkladanej štúdie je prispieť do prebiehajúcej diskusie ohľadom možnosti regulácie online platforiem a systémov umelej inteligencie (systémov AI), ktoré využívajú. Európska komisia predstavila v tomto smere viacero návrhov právnych úprav. Ako uvádzame v časti 1 dokumentu (*Úvod*) našim cieľom bolo analyzovať technologické riešenia online platforiem vo vzťahu k šíreniu dezinformácií a následne posúdiť efektívnosť inštitútov týchto návrhov, najmä návrhu **nariadenia o umelej inteligencii** a **návrhu nariadenia o digitálnych službách**. Konkrétne sme sa zamerali na výskumnú otázku:

“Predstavujú navrhované právne úpravy dostatočne efektívny nástroj pre boj s dezinformáciami v kontexte technologických riešení, ktoré online platformy využívajú?”

V časti 2 dokumentu (*Dezinformácie a ich šírenie v online platformách*) uvádzame, že hoci v právnom poriadku EÚ nemá termín dezinformácia stanovenú legálnu definíciu, viaceré nezáväzné časti právnych aktov, dôvodových správ, štúdií, stratégií a odporúčaní sa zhodujú, že za dezinformáciu môžeme považovať „overiteľne nepravdivú alebo zavádzajúcu informáciu, ktorá je vytvorená, prezentovaná a šírená na účely hospodárskeho zisku alebo zámerného zavádzania verejnosti, a môže poškodiť verejný záujem.“¹

Online platformy a obzvlášť veľmi veľké online platformy zásadne zmenili spôsob, akým ľudia získavajú informácie, a pre mnohých sa stali (hlavným) zdrojom spravodajstva. Poslaním online platforiem pritom nie je vyváženosť a objektívne informovanie svojich užívateľov, ale sú založené na modeli **ekonomiky pozornosti**. Ich cieľom je preto získať čo najviac pozornosti užívateľov za účelom zobrazenia online reklamy, ktorá pre online platformy predstavuje hlavný

¹ Pozri napr.: Európska komisia. OZNÁMENIE KOMISIE EURÓPSKEMU PARLAMENTU, RADE, EURÓPSKEMU HOSPODÁRSKEMU A SOCIÁLNEMU VÝBORU A VÝBORU REGIÓNOV Boj proti dezinformáciám na internete: európsky prístup. COM/2018/236 final. Dostupné na: <https://eur-lex.europa.eu/legal-content/SK/TXT/?uri=CELEX%3A52018DC0236>.

zdroj príjmov. Využívajú na to atraktívne užívateľské rozhrania s prvkami návykovosti a **metódy personalizovaného odporúčania** za účelom poskytnutia pútavého obsahu a relevantnej online reklamy.

Odporúčacie systémy sa snažia znižovať informačné zahltenie užívateľov prostredníctvom filtrovania relevantných informácií. Existuje viacero prístupov odporúčania; personalizované prístupy však vždy musia mať k dispozícii nejakú informáciu o preferenciách užívateľov. Uviedli sme viacero možných dopadov použitia týchto systémov vo všeobecnosti, ako aj konkrétne v kontexte online platforiem.

Ako kľúčové problémy sme identifikovali (ne)transparentnosť generovaných odporúčaní a súčasné nastavenie funkcie užitočnosti odporúčacích systémov v procese tréningu modelov umelej inteligencie. Hoci uspokojivé vysvetlenie, prečo bolo užívateľovi niečo odporúčené, nie je v súčasnosti vždy technicky možné poskytnúť (ide o otvorený výskumný problém), minimálnou požiadavkou je transparentnosť na úrovni vstupov použitých odporúčacím systémom, ako aj na úrovni zvolenej funkcie užitočnosti a nastavenia jej váh, keďže tieto zásadne vplývajú na výstupné odporúčania. Ako ďalšie problémy sme identifikovali uzatváranie užívateľov do informačných bublín, skreslenia a férovosť pri odporúčaní a zbieranie spätnej väzby a odvodzovanie preferencií užívateľa v súvislosti s ochranou súkromia.

V časti 3 dokumentu (*Vybrané právne úpravy navrhované Európskou komisiou vo vzťahu k šíreniu dezinformácií a správaniu sa online platforiem*) sme identifikovali množinu základných práv a slobôd, ktoré môžu byť potenciálne ovplyvnené pri legislatívnom uchopení nástrojov pre boj so šírením dezinformácií na online platformách. Z pohľadu užívateľov to môžu byť predovšetkým zásahy do slobody prejavu, práva na informácie, ochrany súkromia a osobných údajov, či práva na spravodlivý súdny proces. Legislatívne nástroje ale môžu ovplyvniť aj samotné online platformy, a to primárne z hľadiska práva podnikateľ, či ochrany vlastníckeho práva v podobe ochrany práv duševného vlastníctva.

Vzhľadom na deklarované ciele sme sa zamerali na analýzu návrhov novej legislatívy Európskej komisie, ktorej predmetom je umelá inteligencia (návrh nariadenia o umelej inteligencii) a digitálne služby vrátane online platforiem (návrh nariadenia o digitálnych službách). Zamerali sme sa na legislatívne nástroje v týchto dvoch návrhoch, ktoré môžu šírenie dezinformácií v online priestore výraznejšie ovplyvniť.

Návrh nariadenia o umelej inteligencii je koncipovaný ako horizontálna regulácia založená na skúmaní rizika. Systémy umelej inteligencie delí na štyri typy podľa miery (potenciálneho) rizika pre práva a slobody jednotlivcov. V kontexte boja s dezinformáciami bude kľúčové, ako predmetné nariadenie zadefinuje zakázané „manipulatívne praktiky,“ nakoľko **podľa súčasného znenia návrhu by podpora šírenia dezinformácií zo strany online platforiem nemala pod zakázanú praktiku spadať**. Zároveň je nutné poznamenať, že nastavenie klasifikácie systémov AI neumožňuje aktivity online platforiem zaradiť do niektorej z oblastí pre využívanie systémov AI vysokého rizika. Napriek tomu návrh nariadenia upravuje niekoľko

zaujímavých inštitútov týkajúcich sa posúdenia vhodnosti či správy údajov, ktoré môžu pri boji so šírením dezinformácií pomôcť.

Návrh nariadenia o digitálnych službách definuje množiny entít, na ktoré sa bude vzťahovať. V súvislosti so šírením dezinformácií za kľúčové považujeme **požiadavky týkajúce sa online platforiem a veľmi veľkých online platforiem**. Drvivá väčšina požiadaviek v návrhu nariadenia o digitálnych službách sa týka nelegálneho obsahu, kde nie vždy budú dezinformácie spadať. Jednou z mála výnimiek je inštitút posúdenia rizík a prijatie nadväzujúcich opatrení pre veľmi veľké online platformy. Za zaujímavé taktiež považujeme inštitúty týkajúce sa vykonávania externých auditov, transparentnosti reklamy a odporúčacích systémov či povinnosť podávať správu o transparentnosti. Systém zodpovednosti za cudzí obsah sa zásadne pre online platformy nemení, podrobnejšie je však upravené moderovanie obsahu a presný proces a práva užívateľov pri odstraňovaní alebo blokovaní nezákonného obsahu zo strany online platforiem.

Ak vezmeme do úvahy model fungovania online platforiem, metódy odporúčania z technického hľadiska a špecifiká skúmaných návrhov legislatívy z dielne Európskej komisie, máme za to, že **skúmané návrhy právnych aktov je možné vylepšiť takým spôsobom, aby poskytovali účinnejšie nástroje pre boj so šírením dezinformácií**, čomu sme sa venovali v časti 4 (*Diskusia a návrhy riešení*).

Z hľadiska všeobecných pripomienok **považujeme za kľúčové, aby oblasť ekonomiky pozornosti bola považovaná za oblasť systémov AI vysokého rizika**, aby využívanie algoritmov sociálnymi médiami „neušlo“ legislatívnym požiadavkám prezentovaných v návrhu nariadenia o umelej inteligencii.

Zároveň rozumieme limitom právnej úpravy škodlivého obsahu z hľadiska zásahov do slobody prejavu a priznaniu príliš veľkej moci online platformám. Na druhej strane, legislatíva môže poskytovať široké penzum nástrojov, ktoré majú potenciál limitovať šírenie dezinformácií na online platformách bez priamej regulácie škodlivého obsahu. Ako príklady zavedenia takýchto „nepriamych“ pravidiel odporúčame:

- Požiadavky na transparentnosť a užívateľské voľby pri odporúčacích systémoch;
- Označovanie neovereného alebo neoveriteľného obsahu (*content labelling*);
- Zákaz propagovania určitého obsahu alebo tém;
- Obmedzenie používania určitých metód pri citlivom obsahu (napríklad botov alebo mikro-targetingu pri politických reklamách).

Za absolútne zásadné považujeme výkon externých auditov. DSA síce takýto mechanizmus priamo upravuje, ale s obavami sledujeme jeho limity týkajúce sa nedostatku jasných pravidiel pre určenie vhodných entít pre vykonanie externých auditov, nedostatočnej záväznosti výsledkov takýchto auditov a implementáciu ich záverov. Zároveň, prístup externých audítorov by nemal byť obmedzovaný z hľadiska ochrany práv online platformy ako obchodné tajomstvo alebo tretích strán v podobe ochrany osobných údajov. Z technického hľadiska odporúčame ako vhodnú formu auditu tzv. bábkový audit, avšak vyššie uvedené pripomienky je nutné

nastaviť takým spôsobom, aby využívanie botov na účely externého auditu bolo z hľadiska legislatívy a podmienok využívania online platforiem možné. Zvýrazňujeme taktiež rolu vedeckého výskumu a prístup k údajom zo strany preverených výskumníkov.

Dôležitou oblasťou pri boji s dezinformáciami je aj transparentnosť. V tomto smere vítame návrhy prezentované v návrhu nariadenia o digitálnych službách týkajúce sa transparentnosti reklamy a odporúčacích systémov či verejne dostupných správ. Tieto mechanizmy ale musia zabezpečovať, aby ich výsledky boli zmysluplné a ukazovali odbornej a laickej verejnosti skutočný stav. **V súvislosti s transparentnosťou odporúčacích systémov upriamujeme pozornosť na to, aby DSA zakotvila v akejkoľvek forme povinný opt-in pre užívateľov pri prvom kontakte s online platformou.** Odporúčanie by sa napríklad mohlo skladať z dvoch úrovní: nepersonalizovanej a personalizovanej, pričom užívateľ by si mohol zapnúť napr. len tú nepersonalizovanú, teda takú, ktorá pri odporúčaní neberie do úvahy správanie užívateľa, ktoré je zdrojom pre odhad preferencií. Zároveň by uchovávanie logov pre odporúčacie systémy malo byť právne zakotvené v DSA pre prípady, ktoré by potenciálne AIA nepokrývala.

Správy o transparentnosti by mali byť doplnené o riešenie konkrétnych prípadov, aby bolo jasné, ako sa online platforma v konkrétnych situáciách správa a aké opatrenia na zmierňovanie dopadov prijíma. Zároveň by štatistické ukazovatele v týchto správach mali byť rozčlenené podľa členských štátov, aby bolo možné odhaliť prípadné rozdiely v šírení dezinformačného obsahu v jednotlivých členských štátoch, ako aj v prístupe platforiem a orgánov členských štátov a v prijatých opatreniach.

Nemenej dôležité je, aby sociálne médiá pravidelne vyhodnocovali nielen právne, ale aj etické a sociálne riziká. Považujeme za zásadné, aby platformy dokázali vyčleniť čas a zdroje na kontinuálne vyhodnocovanie možných dopadov z pohľadu dotknutých morálnych hodnôt a princípov počas celého cyklu od dizajnu nových funkcionalít až po ich nasadenie. **Posúdenie etických rizík by sa podľa nás malo považovať za záväznú súčasť posudzovania zhody navrhovaného pre poskytovateľov systémov AI v nariadení Európskej komisie o umelej inteligencii.**

Obsah

1. Úvod.....	1
2. Dezinformácie a ich šírenie na online platformách.....	5
2.1 Model fungovania online platforiem	8
2.2 Metódy odporúčania	13
2.2.1 Personalizácia a adaptácia	14
2.2.2 Modelovanie užívateľa.....	14
2.2.3 Personalizované odporúčanie.....	15
2.2.4 Vyhodnocovanie úspešnosti	17
2.2.5 Typické problémy odporúčacích systémov	19
2.2.6 Príklady použitia na online platformách.....	20
2.3 Dopady použitia odporúčacích systémov	23
3. Vybrané právne úpravy navrhované Európskou komisiou vo vzťahu k šíreniu dezinformácií a správaniu sa online platforiem.....	27
3.1 Návrh nariadenia o umelej inteligencii (AIA)	31
3.1.1 Všeobecná charakteristika	31
3.1.2 Boj s dezinformáciami.....	33
3.2 Návrh nariadenia o digitálnych službách (DSA)	36
3.2.1 Všeobecná charakteristika	36
3.2.2 Boj s dezinformáciami.....	39
4. Diskusia a návrhy riešení	49
4.1 Všeobecné pripomienky k regulácii	49
4.1.1 Ekonomika pozornosti ako oblasť systémov AI vysokého rizika	49
4.1.2 Regulácia nezákonného a škodlivého obsahu	50
4.2 Audity a nezávislá kontrola.....	51
4.2.1 Technologické aspekty vykonávania auditov	51
4.2.2 Legislatíva a odporúčania	56
4.2.3 Pozícia vedeckého výskumu	56
4.3 Požiadavky na transparentnosť.....	57
4.4 Správy o činnosti platforiem.....	59
4.5 Posudzovanie z pohľadu etiky.....	59
4.6 Závěry.....	62

Zoznam použitých skratiek

AI znamená umelá inteligencia.

EÚ znamená Európska únia.

EK znamená Európska komisia.

GDPR znamená nariadenie Európskeho parlamentu a Rady (EÚ) 2016/679 z 27. apríla 2016 o ochrane fyzických osôb pri spracúvaní osobných údajov a o voľnom pohybe takýchto údajov, ktorým sa zrušuje smernica 95/46/ES (všeobecné nariadenie o ochrane údajov).

AIA znamená návrh nariadenia Európskeho parlamentu a rady, ktorým sa stanovujú harmonizované pravidlá v oblasti umelej inteligencie (Akt o umelej inteligencii) a menia niektoré legislatívne akty Únie.

DSA znamená návrh nariadenia Európskeho parlamentu a rady o jednotnom trhu s digitálnymi službami (akt o digitálnych službách) a o zmene smernice 2000/31/ES.

Smernica o elektronickom obchode znamená smernica 2000/31/ES Európskeho parlamentu a Rady z 8. júna 2000 o určitých právnych aspektoch služieb informačnej spoločnosti na vnútornom trhu, najmä o elektronickom obchode (smernica o elektronickom obchode).

Charta základných práv znamená Charta základných práv Európskej únie.

Dohovor znamená Dohovor o ochrane ľudských práv a základných slobôd.

Ústava SR znamená 460/1992 Zb. Ústava Slovenskej republiky.

1. Úvod

Šírenie informácií verejnosti masovokomunikačnými prostriedkami bolo v minulosti doménou tlačových správ a neskôr televízneho vysielania. S nástupom nových technológií a hlavne internetu sa ale situácia v ostatnom období radikálne zmenila. Šíriť akékoľvek informácie je v dnešnej dobe mimoriadne jednoduché, pričom hlavnou platformou sa stávajú populárne sociálne médiá s miliónmi registrovaných užívateľov. Každý z nás si dnes môže založiť blog alebo fanúšikovskú stránku na sociálnej sieti a prezentovať svoje výstupy širokej verejnosti. Nové technológie nemajú na šírenie informácií iba pozitívny efekt v podobe väčšej dostupnosti informácií, ale zároveň zvyrazňujú jeden zo súčasných problémov informačného ekosystému - dezinformácie. Dezinformácie sú veľmi atraktívnym obsahom, nakoľko poskytujú zdanlivo jednoduché odpovede na komplexné otázky. Často fungujú v rámci konkrétneho príbehu, ktorý je pre ľudí zaujímavý.²

Šírenie dezinformácií predstavuje hybridnú hrozbu, ktorá má za následok polarizovanie spoločnosti a môže viesť až k vnútornej destabilizácii demokratických inštitúcií.³ Aj prebiehajúca pandémia COVID-19 ukazuje, že dôvera v dezinformačné naratívy môže viesť k vážnym ujám na zdraví, prípadne strate na životoch. Podľa prieskumu GLOBSEC, ktorý analyzoval postoj obyvateľov krajín strednej a východnej Európy k dezinformáciám, na Slovensku až 56% ľudí dôveruje takémuto obsahu.⁴

Sociálne médiá a ich činnosť predstavujú významnú premennú pri šírení dezinformácií, nakoľko disponujú nástrojmi, ktoré môžu byť využité na informovanie veľkej časti publika. Dezinformácie sa jednoducho môžu dostať k ovplyvniteľným častiam obyvateľstva, ktoré nemajú jednoznačný postoj na určitú tému. Ako ukazuje prieskum Ministerstva obrany Slovenskej republiky, obsah dezinformačných webov sa k užívateľom vo veľmi veľkej miere dostáva práve cez sociálnu sieť Facebook.⁵ Problém zvyrazňujú aj nové techniky a metódy na šírenie dezinformácií ako zakladanie troch fariem, Potemkinove osoby,⁶ využívanie umelej inteligencie, "deepfakes", presun do privátnych častí sociálnych médií a šírenie zvukových správ.⁷ Na strane druhej, systémy umelej inteligencie môžu byť využité aj ako nástroj na boj s dezinformáciami.⁸

² Bližšie pozri napríklad Európsky Parlament. Disinformation and propaganda: impact on the functioning of the rule of law and democratic processes in the EU and its Member States - 2021 update. Dostupné na: [https://www.europarl.europa.eu/thinktank/en/document.html?reference=EXPO_STU\(2021\)653633](https://www.europarl.europa.eu/thinktank/en/document.html?reference=EXPO_STU(2021)653633).

³ Ministerstvo Obrany Slovenskej republiky. Cena za (dez)informácie. Výnosy z reklám webov s problematickým obsahom. Dostupné na: https://www.mosr.sk/data/files/4364_2021-k-01-cena-za-dezinformacie-recenzovane.pdf.

⁴ GLOBSEC. Voices of Central and Eastern Europe, 2020. Dostupné na: <https://www.globsec.org/publications/voices-of-central-and-eastern-europe/>.

⁵ Ministerstvo Obrany Slovenskej republiky. Cena za (dez)informácie. Výnosy z reklám webov s problematickým obsahom. Dostupné na: https://www.mosr.sk/data/files/4364_2021-k-01-cena-za-dezinformacie-recenzovane.pdf.

⁶ Entity, ktoré sa tvária ako kredibilné inštitúcie (napríklad neziskové organizácie) s veľkou základňou naprieč internetom. Zároveň ale do pravdivého obsahu vkladajú dezinformácie. Pozri napr. DIRESTA, R. - GROSSMAN, S. Potemkin Pages & Personas: Assessing GRU Online Operations, 2014-2019. White paper, Stanford Internet Observatory Cyber Policy Center, Stanford, 2019.

⁷ Európsky Parlament. Disinformation and propaganda: impact on the functioning of the rule of law and democratic processes in the EU and its Member States - 2021 update. Dostupné na: [https://www.europarl.europa.eu/thinktank/en/document.html?reference=EXPO_STU\(2021\)653633](https://www.europarl.europa.eu/thinktank/en/document.html?reference=EXPO_STU(2021)653633).

⁸ Bližšie pozri napr. GUO, B. - DING, Y. - YAO, L. - LIANG, Y. - YU, Z. The Future of False Information Detection on Social Media: New Perspectives and Trends. ACM Computing Surveys, 53(4), 1–36, 2020. Dostupné na: <https://doi.org/10.1145/3393880>; alebo výsledky projektov financovaných zo zdrojov EÚ, napr. [WeVerify](#), [Provenance](#), alebo [EDMO](#).

Rola online platforiem a hlavne sociálnych médií je pri šírení dezinformácií značná. Na jednej strane predstavujú ideálny priestor pre šírenie akýchkoľvek informácií a prispievajú k napĺňaniu ideálov informačnej spoločnosti. Na strane druhej ale svojou činnosťou či už úmyselne alebo neúmyselne⁹ prispievajú k ich šíreniu prostredníctvom zvýhodňovania obsahu, ktorý má potenciálne väčší dosah (ide o obsah spravidla šokujúcejší, kontroverzejší či vyvolávajúci emócie). Prispievajú k tomu aj nastavenia parametrov systémov umelej inteligencie využívané na personalizáciu obsahu zobrazeného jednotlivým užívateľom. Zároveň odborná, ale aj laická verejnosť často kritizuje nepružnosť sociálnych médií pri moderovaní obsahu a (ne)odstraňovaní problematického obsahu. Regulácie sociálnych médií v podobe nastavenia opatrení pre boj s dezinformáciami sú preto v súčasnosti diskutovanou témou a vzhľadom na globálnosť problematiky je prirodzené, že Európska únia chce zaujať v tejto oblasti regulácie pozíciu svetového lídra.

Viaceré štáty v EÚ sa ale rozhodli šírenie dezinformácií už právne upraviť vo svojich národných právnych poriadkoch. Francúzsko v roku 2018 predstavilo zákon o boji proti manipulovaniu s informáciami.¹⁰ Predmetný zákon umožňuje v čase tri mesiace pred voľbami francúzskym súdom vydať príkaz na odstránenie obsahu, ktorý obsahuje nepresné alebo zavádzajúce informácie šírené úmyselne, umelo alebo automatizovane vo veľkom rozsahu s cieľom ovplyvniť dôveryhodnosť volieb. Zároveň je šírenie falošných správ v súvislosti s ovplyvňovaním výsledkov volieb vo Francúzsku trestným činom.¹¹ Trestný zákon Malty taktiež považuje šírenie falošných správ za trestný čin.¹² Podobné zákazy obsahuje aj právny poriadok Litvy, ktorý priamo zakazuje šírenie dezinformácií,¹³ či trestný zákon Grécka.¹⁴ Trestnoprávne dôsledky šírenia nepravdivých správ obsahujú aj trestné kódexy Rakúska¹⁵ a Poľska.¹⁶ Viaceré štáty prijali podobné opatrenia ako prevenciu v rámci pandémie COVID-19. Spomenúť možno Rumunsko¹⁷ a Maďarsko.¹⁸ V súčasnosti aj Slovenská republika predstavila novelizáciu trestného zákona, ktorá by mala zaviesť trestný čin šírenia dezinformácií.

Napriek tomu, že aj v prostredí právnych poriadkov EÚ existujú zákony, ktoré šírenie dezinformácií trestajú rôznym spôsobom, je potrebné zvýrazniť, že európska právna kultúra je budovaná na silnom rešpekte k základným ľudským právam a slobodám. Boj proti šíreniu dezinformácií významne naráža na slobodu prejavu a právo na informácie. Ako sa uvádza

⁹ The Wall Street Journal. The Facebook Files. Dostupné na: <https://www.wsj.com/articles/the-facebook-files-11631713039>.

¹⁰ Loi 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information [Law 2018-1202 of December 22, 2018 on the fight against the manipulation of information], JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE]. Citované z HOBOKEN VAN, J. - Ó FATHAIGH, R. Regulating Disinformation in Europe: Implications for Speech and Privacy. UC Irvine Journal of International, Transnational, and Comparative Law. Volume 6 Symposium: The Transnational Legal Ordering of Privacy and Speech. Article 3, 2021.

¹¹ Code électoral [Electoral Code] art. L97. Citované z HOBOKEN VAN, J. - Ó FATHAIGH, R. Regulating Disinformation in Europe: Implications for Speech and Privacy. UC Irvine Journal of International, Transnational, and Comparative Law. Volume 6 Symposium: The Transnational Legal Ordering of Privacy and Speech. Article 3, 2021.

¹² CRIMINAL CODE, § 82 (amended by the Media and Defamation Act, 2018).

¹³ Law on the Provision of Information to the Public, No I-1418, § 2 (13) (1996), amended by No XII-2239 of Dec. 23 2015.

¹⁴ POINIKOS KODIKAS [P.K] [CRIMINAL CODE] 4619:191.

¹⁵ Strafgesetzbuch [StGB] [Penal Code] § 264 [Verbreitung falscher Nachrichten bei einer Wahl oder Volksabstimmung] [Dissemination of false news in an election or referendum].

¹⁶ Ordynacja wyborcza do rad gmin, rad powiatów i sejmików województw [Law of 16 July 1998 on Elections to Municipalities, District Councils and Regional Assemblies] (Dz.U. 1998 Nr 95 poz. 602), § 72.

¹⁷ Decret semnat de Președintele României, domnul Klaus Iohannis, privind instituirea stării de urgență pe teritoriul României, 16 martie 2020 [Decree Signed by the President of Romania, Mr. Klaus Iohannis, Regarding the Establishment of the State of Emergency on the Romanian Territory, Mar. 16, 2020] § 54.

¹⁸ 2020. évi XII (Act XII of 2020 on the containment of coronavirus), § 10(2), amending the Criminal Code, § 337.

v medzinárodných dokumentoch k danej problematike, všeobecné zákazy pre šírenie informácií by nemali byť založené na vágnych myšlienkach a definíciách ako falošné správy alebo dezinformácie.¹⁹ Podobne sa vyjadril aj špeciálny spravodajca Organizácie spojených národov, ktorý varoval pred právomocami orgánov verejnej moci vymazať alebo blokovat obsah podľa vlastného konceptu pravdivosti.²⁰ Takýto prístup potvrdzuje aj judikatúra Európskeho súdu pre ľudské práva, kde v kontexte šírenia dezinformácií súd potvrdil legalitu obmedzenia práva na informácie iba vo veľmi zjavných a potvrdených prípadoch, ako je popieranie holokaustu.²¹ Aj z týchto dôvodov je prístup Európskej komisie k regulácii boja proti dezinformáciám veľmi opatrný.

Vzhľadom na vyššie uvedené dôvody a globálnosť problematiky sa v predkladanej štúdii zameriame predovšetkým na právnu úpravu boja proti dezinformáciám na úrovni EÚ.

Primárne ciele štúdie sú:

- a) analyzovať technologické riešenia online platforiem vo vzťahu k šíreniu dezinformácií a
- b) následne posúdiť efektívnosť inštitútov vybraných právnych úprav navrhovaných Európskou komisiou, najmä návrhu nariadenia o umelej inteligencii²² a návrhu nariadenia o digitálnych službách.²³

Vyššie uvedené ciele sú reflektované prostredníctvom výskumnej otázky:

“Predstavujú navrhované právne úpravy dostatočne efektívny nástroj pre boj s dezinformáciami v kontexte technologických riešení, ktoré online platformy využívajú?”

Napriek tomu, že počas vypracovania štúdie využívame prísny a na mieru šitý vedecký prístup s cieľom maximalizovať platnosť a spoľahlivosť zistení, uznávame limity predkladanej štúdie. Tieto limity predovšetkým predstavujú: minimálna transparentnosť a nedostatok zdrojových informácií o technickom fungovaní online platforiem, časový rámec vypracovania štúdie (október a november 2021) a aktuálnosť návrhov právnych úprav a platnej legislatívy; a zameranie sa špecificky na otázky regulácie online platforiem v podobe návrhu nariadenia o digitálnych službách a návrhu nariadenia o umelej inteligencii.

S ohľadom na naplnenie cieľov a odpovede na výskumnú otázku predkladanú štúdiu delíme na tri ťažiskové časti. V časti 2 sa štúdia venuje pojmu dezinformácie a vysvetleniu fungovania

¹⁹ Joint Declaration on Freedom of Expression and “Fake News”, Disinformation and Propaganda, FOM.GAL/3/17, (Mar. 3, 2017). Dostupné na: <https://www.osce.org/files/f/documents/6/8/302796.pdf>.

²⁰ KAYE, D (Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression). Disease Pandemics and the Freedom of Opinion and Expression, U.N. Doc. A/HRC/44/49 (Apr. 23, 2020), s. 13.

²¹ Rozhodnutie Európskeho súdu pre ľudské práva vo veci Garaudy v France. Sťažnosť č. 65831/01, 24 Jún 2003.

²² Návrh nariadenia Európskeho parlamentu a rady, ktorým sa stanovujú harmonizované pravidlá v oblasti umelej inteligencie (Akt o umelej inteligencii) a menia niektoré legislatívne akty Únie. Ďalej v texte označujeme ako návrh nariadenia o umelej inteligencii alebo používame skratku AIA (z angl. Artificial Intelligence Act). Dostupné na: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>.

²³ Návrh nariadenia Európskeho parlamentu a rady o jednotnom trhu s digitálnymi službami (akt o digitálnych službách) a o zmene smernice 2000/31/ES. Ďalej v texte označujeme ako návrh nariadenia o digitálnych službách alebo používame skratku DSA (z angl. Digital Service Act) v rovnakom význame. Dostupné na: <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital-services>.

online platforiem z pohľadu šírenia dezinformácií. Pojem online platforma používame v súlade s navrhovaným legislatívnym vymedzením v návrhu nariadenia o digitálnych službách. Osobitnú pozornosť venujeme modelom fungovania online platforiem a metódam odporúčania z technického hľadiska.

Časť 3 sa zameriava na legislatívu a právne nástroje boja s dezinformáciami. Predmetná časť začína exkurzom do základných ľudských práv a slobôd, ktoré môžu byť v súvislosti so šírením dezinformácií ohrozené. Následne sa osobitne venujeme konkrétnym inštitútom vybraných právnych úprav navrhovaných Európskou komisiou, najmä návrhu nariadenia o umelej inteligencii a návrhu nariadenia o digitálnych službách.

Časť 4 predstavuje diskusiu a navrhované opatrenia, ktoré vychádzajú z analýzy technického fungovania online platforiem a nedostatkov navrhovaných právnych úprav.

2. Dezinformácie a ich šírenie na online platformách

Dezinformácie predstavujú jednu z najväčších výziev pre súčasné iniciatívy pri tvorbe pravidiel s cieľom regulovať obsah a zodpovednosť na internete. Nie len pandémie COVID-19 a šírenie informácií o vakcinácii sú toho aktuálnym dôkazom. EÚ ako "regulačná supervelmoc" s výrazným dosahom na právne poriadky tretích krajín²⁴ sa postavila do čela regulácie, ktorá by mala prispieť k boju s dezinformáciami.

Samotný termín dezinformácia nemá v právnom poriadku EÚ stanovenú legálnu definíciu. Pri definovaní predmetného pojmu môžeme vychádzať iba z nezáväzných častí právnych aktov, dôvodových správ, štúdií, stratégií a odporúčaní.

Oznámenie Európskej komisie týkajúce sa európskeho prístupu k boju s dezinformáciami uvádza:

"Za dezinformáciu sa považuje overiteľne nepravdivá alebo zavádzajúca informácia, ktoré je vytvorená, prezentovaná a šírená na účely hospodárskeho zisku alebo zámerného zavádzania verejnosti, a môže poškodiť verejný záujem. Poškodenie verejného záujmu zahŕňa ohrozenie demokratických procesov a procesy tvorby politiky, ako aj verejných statkov, ako napr. ochrany zdravia občanov EÚ, životného prostredia alebo bezpečnosti. Medzi dezinformácie nepatria chyby v spravodajstve, satira a paródie, ani správy a komentáre naklonené jednej strane, ktoré sú takto zreteľne označené."²⁵

Túto definíciu preberá aj Akčný plán Európskej komisie pre boj s dezinformáciami.²⁶

Podobne definuje pojem dezinformácia aj Správa expertnej skupiny na vysokej úrovni o falošných správach a dezinformáciách v online prostredí. Za dezinformáciu správa považuje nepravdivé, nepresné alebo zavádzajúce informácie navrhnuté, prezentované a propagované s cieľom úmyselne spôsobiť spoločenskú škodu alebo za účelom zisku.²⁷

Kódex nakladania s dezinformáciami definuje dezinformácie pozitívne a negatívne. Z pozitívneho hľadiska definícia dezinformácie kopíruje vyššie uvedené vymedzenie:²⁸

"overiteľne nepravdivá alebo zavádzajúca informácia", ktorá je súhrnne

²⁴ BRADFORD, A. The Brussels Effect. How the European Union Rules the World. Oxford University Press Inc, 2020.

²⁵ Európska komisia. OZNÁMENIE KOMISIE EURÓPSKEMU PARLAMENTU, RADE, EURÓPSKEMU HOSPODÁRSKEMU A SOCIÁLNEMU VÝBORU A VÝBORU REGIÓNOV Boj proti dezinformáciám na internete: európsky prístup. COM/2018/236 final. Dostupné na: <https://eur-lex.europa.eu/legal-content/SK/TXT/?uri=CELEX%3A52018DC0236>.

²⁶ Európska komisia. SPOLOČNÉ OZNÁMENIE EURÓPSKEMU PARLAMENTU, EURÓPSKEJ RADE, RADE, EURÓPSKEMU HOSPODÁRSKEMU A SOCIÁLNEMU VÝBORU A VÝBORU REGIÓNOV Akčný plán proti dezinformáciám. JOIN/2018/36 final. Dostupné na: <https://eur-lex.europa.eu/legal-content/SK/TXT/?uri=CELEX%3A52018JC0036>.

²⁷ A multi-dimensional approach to disinformation. Report of the independent High level Group on fake news and online disinformation, 2018, s. 10. Originál: "...false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit." Dostupné na: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=50271.

²⁸ Kódex nakladania s dezinformáciami. 2018. Dostupné na: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=59125.

- a) „vytvorená, prezentovaná a šírená na účely hospodárskeho zisku alebo zámerného zavádzania verejnosti“, a
- b) „môže poškodiť verejný záujem“, keďže jej ciele zahŕňajú „ohrozenie demokratických procesov a procesy tvorby politiky, ako aj verejných statkov, ako napr. ochrany zdravia občanov EÚ, životného prostredia alebo bezpečnosti.“

Negatívna definícia dezinformácie z tohto pojmu vylučuje „klamlivú reklamu, chyby v spravodajstve, satiru a paródie ani správy a komentáre naklonené jednej strane, ktoré sú takto zreteľne označené, a nemá vplyv na platné právne predpisy, samoregulačné kódexy v oblasti reklamy a normy týkajúce sa klamlivej reklamy.“

DSA (návrh nariadenia o digitálnych službách) vo svojej recitálovej časti oddeľuje pojmy dezinformácia od ilegálneho obsahu²⁹ či iných manipulatívnych a nekalých činností.³⁰

Na úrovni Slovenskej republiky, podobne ako pri práve EÚ, legálna definícia pojmu dezinformácia neexistuje. Pri výklade daného pojmu je však možné oprieť sa o strategické dokumenty SR. Koordinovaný mechanizmus odolnosti Slovenskej republiky voči informačným operáciám obsahuje nasledovnú definíciu:

„Dezinformácia - označuje nepravdivú alebo zmanipulovanú informáciu, ktorá je šírená zámerne s cieľom zavádzať a uškodiť. Dezinformácie môžu mať podobu nepravdivého alebo zmanipulovaného textu, obrázku, videa alebo zvuku, pričom môžu byť použité na podporu konšpirácií, šírenie pochybností a diskreditáciu pravdivých informácií či jednotlivcov a organizácií. Aj pravdivú informáciu môžeme považovať za dezinformáciu, ak je podaná manipulatívnym spôsobom. Medzi dezinformácie nepatria neúmyselné chyby v spravodajstve, satira, paródia ani správy a komentáre naklonené jednej strane, ktoré sú takto zreteľne označené.“³¹

Bezpečnostná stratégia SR³² či Koncepcia Slovenskej republiky pre boj proti hybridným hrozbám³³ síce šírenie dezinformácií výslovne spomínajú na viacerých miestach, ale samotný termín nedefinujú.

²⁹ DSA, Recitál 71.

³⁰ DSA, Recitál 68.

³¹ Koordinovaný mechanizmus odolnosti Slovenskej republiky voči informačným operáciám. Dostupné na slov-lex.sk.

³² Bezpečnostná stratégia Slovenskej republiky. Dostupné na: <https://www.nbu.gov.sk/wp-content/uploads/urad/Bezpecnostna-strategia-SR-2021.pdf>.

³³ Koncepcia pre boj Slovenskej republiky proti hybridným hrozbám. Dostupné na: <https://www.nbu.gov.sk/wp-content/uploads/PHHD/Koncepcia-boja-SR-proti-hybridnym-hrozbam.pdf>.

Trestný zákon Slovenskej republiky

Zaujímavosťou je, že slovenský Trestný zákon obsahuje viacero pojmov, ktoré s dezinformáciami súvisia, ale priamo ich nedefinuje.

Trestný čin ohrozenia bezpečnosti vzdušného dopravného prostriedku a lode upravuje skutkovú podstatu, v zmysle ktorej „kto oznámi **nepravdivú informáciu**, ktorá **môže ohroziť bezpečnosť alebo prevádzku vzdušného dopravného prostriedku za letu alebo lode za plavby**, potrestá sa odňatím slobody až na tri roky.“

S pojmom nepravdivá informácia operuje aj trestný čin teroristického útoku: „kto v úmysle poškodiť ústavné zriadenie alebo obranyschopnosť štátu, narušiť alebo zničiť základnú politickú, hospodársku alebo spoločenskú štruktúru štátu alebo medzinárodnej organizácie, závažným spôsobom zastrašiť obyvateľstvo alebo donútiť vládu štátu alebo iný orgán verejnej moci alebo medzinárodnej organizácie, aby niečo konala, opomenula alebo strpela... zmocní sa lietadla, lode, iného prostriedku osobnej dopravy alebo...**oznámi nepravdivú informáciu**, čím ohrozí život alebo zdravie ľudí, bezpečnosť takého dopravného prostriedku.“

Ďalším príkladom je trestný čin šírenia poplašnej správy: „kto úmyselne **spôsobí nebezpečenstvo vážneho znepokojenia aspoň časti obyvateľstva nejakého miesta** tým, že **rozširuje poplašnú správu, ktorá je nepravdivá**, alebo sa dopustí iného obdobného konania spôsobilého vyvolať také nebezpečenstvo, potrestá sa odňatím slobody až na dva roky.“

Trestný čin ohovárania zasa operuje s termínom nepravdivý údaj: „kto **o inom oznámi nepravdivý údaj**, ktorý je spôsobilý značnou mierou ohroziť jeho vážnosť u spoluobčanov, poškodiť ho v zamestnaní, v podnikaní, narušiť jeho rodinné vzťahy alebo spôsobiť mu inú vážnu ujmu, potrestá sa odňatím slobody až na dva roky.“

Právna literatúra dezinformácie vníma z troch pohľadov

- obsahového,
- účastníkov a
- distribúcie.

Z pohľadu obsahu ide o klasifikovanie informácie, ktorá je nepravdivá alebo zavádzajúca. Práve tento aspekt je výrazne predmetom regulácie. Perspektíva účastníkov odzrkadľuje dezinformácie z pohľadu entity, ktorá ich šíri, napríklad tretie krajiny so záujmom destabilizácie demokratického systému. Tretí pohľad na dezinformácie sa týka šírenia a rôznych techník znásobujúcich efektívnosť distribúcie dezinformácií.³⁴

Okrem pojmu dezinformácia sa v literatúre často používajú ďalšie súvisiace (a niekedy aj čiastočne sa prekrývajúce) pojmy, ako sú falošné informácie, misinformácie, falošné správy, hoaxy a pod. (pre bližšie vysvetlenie vzťahov medzi týmito konceptmi pozri napr. Hrčková a

³⁴ Citované z HOBOKEN VAN, J. - Ó FATHAIGH, R. Regulating Disinformation in Europe: Implications for Speech and Privacy. UC Irvine Journal of International, Transnational, and Comparative Law. Volume 6 Symposium: The Transnational Legal Ordering of Privacy and Speech. Article 3, 2021, s. 13.

kol. (2019)³⁵). Gelfert (2018)³⁶ považuje *falošné správy* (angl. *fake news*) za druh dezinformácií a definuje ich ako "úmyselné vydávanie (typicky) falošných alebo zavádzajúcich tvrdení za *spravodajstvo*, pričom tvrdenia sú *schválne* zavádzajúce." Tento pojem sa začal intenzívne používať v období po roku 2016 v súvislosti s referendom o vystúpení Veľkej Británie z EÚ a prezidentskými voľbami v USA, keď obe kampane boli sprevádzané veľkým množstvom dezinformácií v podobe falošných správ šírených často na veľkých online platformách (Twitter, Facebook).^{37,38}

V súčasnosti sa od používania pojmu falošná správa upúšťa; ako uvádza vyššie citovaná Správa expertnej skupiny na vysokej úrovni o falošných správach a dezinformáciách v online prostredí, je to pre jeho nepresnosť, neúplnosť, ako aj jeho zneužívanie na diskreditáciu politických oponentov; namiesto toho sa prikláňa k pojmu dezinformácia.³⁹ Z podobných dôvodov sa pojmu falošné správy vyhýba aj správa vypracovaná pre Radu Európy, ktorá rozlišuje medzi *misinformáciami* (falošné informácie, ktoré sú zdieľané bez úmyslu ublížiť), *dezinformáciami* (falošné informácie, ktorých úmyslom je spôsobiť škodu) a *malinformáciami* (pravdivé informácie, ktoré sú zdieľané s úmyslom ublížiť, napr. zverejnenie súkromných alebo tajných informácií, ale aj nenávistné prejavy).⁴⁰

Odlíšenie misinformácií a dezinformácií na základe úmyslu toho, kto informáciu zdieľa, je prítomné aj v literatúre z oblasti informatiky (pozri napr. Kumar a Shah (2018)⁴¹). V praxi je však často ťažké úmysel rozlíšiť a pri snahách o moderovanie obsahu alebo detekcii mis-, resp. dezinformácií pomocou automatických metód sa tieto často zamieňajú.

2.1 Model fungovania online platforiem

Online platformy a predovšetkým veľmi veľké online platformy (pre ich definíciu a rozlíšenie podľa DSA pozri sekciu časť 3.2.1) počas svojej relatívne krátkej existencie zmenili, ako ľudia získavajú informácie (v akej podobe, z akých zdrojov, ako s nimi ďalej narábajú a pod.). Ide pritom o celosvetový fenomén; pre ilustráciu, Facebook, ktorý vznikol v roku 2004, má v súčasnosti celosvetovo viac ako 2,89 miliardy aktívnych užívateľov, YouTube (vznikol v roku 2005, v roku 2006 ho odkúpila spoločnosť Google) viac ako 2,29 miliardy užívateľov a WhatsApp (vznikol v roku 2009, v roku 2014 ho odkúpila spoločnosť Facebook) je celosvetovo treťou najpoužívanjšou online platformou s 2 miliardami užívateľov (pozri tabuľku č. 1 pre štatistiky ďalších rozšírených online platforiem).

³⁵ HRČKOVÁ, A. - SRBA, I. - MÓRO, R. - BLAHO, R. - ŠIMKO, J. - NÁVRAT, P. - BIELIKOVÁ, M. Unravelling the basic concepts and intents of misbehavior in post-truth society. *Bibliotecas. Anales de Investigación*; 15(3), 2019, s. 421-428. Dostupné na: <http://revistas.bnjm.cu/index.php/BAI/article/view/109/110>.

³⁶ GELFERT, A. Fake News: A Definition. *Informal Logic*, Vol.38, No.1, 2018, s. 84-117.

³⁷ BASTOS, M. T. - MERCEA, D. The Brexit Botnet and User-Generated Hyperpartisan News, *Social Science Computer Review*, 37(1), 2019, s. 38-54. Dostupné na: <https://doi.org/10.1177/0894439317734157>.

³⁸ BOVET, A. - MAKSE, H.A. Influence of fake news in Twitter during the 2016 US presidential election. *Nat Commun* 10, 7 (2019). Dostupné na: <https://doi.org/10.1038/s41467-018-07761-2>.

³⁹ A multi-dimensional approach to disinformation. Report of the independent High level Group on fake news and online disinformation, 2018. Dostupné na: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=50271.

⁴⁰ WARDLE, C. - DERAKHSHAN, H. Information Disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe report DGI, 2017, 09. Dostupné na: <https://firstdraftnews.org/articles/coe-report/>.

⁴¹ KUMAR, S. - SHAH, N. 2018. False information on web and social media: A survey. *arXiv preprint*. Dostupné na: [arXiv:1804.08559](https://arxiv.org/abs/1804.08559).

Nástup online platforiem, predovšetkým sociálnych médií, zásadným spôsobom zmenil aj prístup ľudí k spravodajstvu. Podľa prieskumu realizovanom Reuters inštitútom⁴² už v roku 2012 v mnohých krajinách ľudia konzumovali spravodajstvo online (napr. v USA 86% respondentov, vo Veľkej Británii 82%, vo Francúzsku 77%, v Nemecku 61%), avšak väčšinou priamo z webových stránok novín alebo vysielateľov (televízií); sociálne médiá a blogy ako zdroj spravodajstva uvádzalo 36% respondentov v USA, ale len 18% v Nemecku a vo Veľkej Británii a 17% vo Francúzsku.

Tabuľka č.1

Štatistiky používania (veľmi veľkých) online platforiem. Online platformy sú zoradené podľa počtu užívateľov celosvetovo, pričom sú zobrazené len platformy rozšírené aj na Slovensku (nie sú zobrazené napr. WeChat na 6. mieste, Sina Weibo na 10. a pod.). Údaje k celosvetovému počtu užívateľov zachytávajú počet aktívnych užívateľov platformy k októbru 2021 s výnimkou platforiem WhatsApp a Facebook Messenger, keďže tieto platformy nepublikovali aktualizované počty užívateľov za posledných 12 mesiacov. Údaje o potenciálnej veľkosti publika na Slovensku (k januáru 2021) vyjadrujú počet ľudí, ktorí môžu byť zasiahnutí reklamou na danej platforme a skutočné počty užívateľov sa od týchto hodnôt môžu preto líšiť. Pre platformy WhatsApp, TikTok a Telegram sa autorom tejto štúdie relevantné údaje pre potenciálnu veľkosť publika na Slovensku nepodarilo získať.

Zdroj údajov: Wikipedia (rok vzniku), Statista.com⁴³ (počet užívateľov celosvetovo), Digital 2021: Slovakia⁴⁴ (potenciálna veľkosť publika na Slovensku).

#	Online platforma	Rok vzniku	Počet užívateľov celosvetovo	Potenciálna veľkosť publika na Slovensku
1.	Facebook	2004	2,89 miliárd	2,70 miliónov
2.	YouTube	2005	2,29 miliárd	4,03 miliónov
3.	WhatsApp	2009	2,00 miliárd	-
4.	Instagram	2010	1,39 miliárd	1.40 miliónov
5.	Facebook Messenger	2011	1,30 miliárd	2.40 miliónov
7.	TikTok	2016	1,00 miliárd	-
11.	Telegram	2013	550 miliónov	-
12.	Snapchat	2011	538 miliónov	465 tisíc
15.	Twitter	2006	463 miliónov	140 tisíc

⁴² NEWMAN, N. Reuters Institute Digital News Report 2012: Tracking the Future of News. Dostupné na: <https://www.digitalnewsreport.org/survey/2012/>.

⁴³ Statista.com. Dostupné na: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>.

⁴⁴ Digital 2021. Slovakia. Report. Dostupné na: <https://datareportal.com/reports/digital-2021-slovakia>.

Naproti tomu v roku 2016 na vzorke 26 krajín už viac ako polovica všetkých respondentov (51%; v EÚ v priemere 46%) uvádzala, že používa sociálne médiá ako zdroj spravodajstva každý týždeň a pre 12% predstavovali hlavný zdroj⁴⁵.

V roku 2021⁴⁶ 66% respondentov z 12 krajín⁴⁷ používalo jedno alebo viaceré sociálne médiá a aplikácie na instantnú komunikáciu za účelom konzumovania, zdieľania a diskutovania spravodajstva. V celom tomto období bol dominantnou platformou Facebook, ktorý však v posledných rokoch oslabil na úkor nastupujúcich sociálnych médií a predovšetkým aplikácií pre instantnú komunikáciu (WhatsApp, Telegram). Aplikácie pre instantnú komunikáciu sú ako zdroj spravodajstva využívané predovšetkým v krajinách Latinskej Ameriky (40% používa WhatsApp) a Afriky (61% používa WhatsApp a 18% Telegram)⁴⁸. Údaje pre Slovensko sú uvedené v tabuľke č. 2.

Tabuľka č. 2

Najpoužívanéjšie online platformy pre získavanie spravodajstva na Slovensku. Sociálne médiá ako zdroj spravodajstva na Slovensku používa celkovo 56% respondentov podľa štúdie Reuters inštitútu z roku 2021.

Zdroj: Reuters Institute, Digital News Report 2021.

Online platforma	Podiel respondentov používajúcich platformu na získavanie spravodajstva	Podiel respondentov používajúcich platformu na ľubovoľný účel
Facebook	55%	72%
YouTube	26%	77%
Facebook Messenger	19%	51%
Instagram	11%	29%
WhatsApp	9%	29%
Viber	5%	20%

Online platformy sú často spájané so šírením dezinformácií. Vo všeobecnosti považovalo dezinformácie za problém 58% respondentov (v krajinách Európy 54%). Špecificky ohľadom dezinformácií o COVID-19 si najviac respondentov robilo starosti v súvislosti s informáciami na Facebooku (28%), na stránkach a v aplikáciách novín (17%), v aplikáciách pre instantnú komunikáciu (15%), vo výsledkoch vyhľadávania (7%) či na YouTube (6%).⁴⁹

⁴⁵ NEWMAN, N. et al. Reuters Institute Digital News Report 2016. Dostupné na: <https://www.digitalnewsreport.org/survey/2016/>.

⁴⁶ NEWMAN, N. et al. Reuters Institute Digital News Report 2021. Dostupné na: <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2021>.

⁴⁷ Tamže. Autori z Reuters inštitútu zvolili 12 krajín (Veľká Británia, USA, Nemecko, Francúzsko, Španielsko, Taliansko, Írsko, Dánsko, Fínsko, Japonsko, Austrália a Brazília), pre ktoré mali údaje dostupné od roku 2014, resp. 2015, aby mohli porovnať zmeny v tomto období.

⁴⁸ Tamže.

⁴⁹ Tamže.

Vzhľadom na to, aký zásadný význam má vyvážené spravodajstvo a slobodný prístup k informáciám pre spoločnosť a demokraciu, je z uvedených štatistík jasné, prečo sa upriamila toľká pozornosť práve na online platformy v čase, keď pozorujeme narastajúcu polarizáciu spoločnosti a šírenie dezinformácií.

Problémom zostáva, že **poslaním online platforiem nie je vyváženosť a objektívne informovanie svojich užívateľov.**

Väčšina ich príjmov (85,9 miliardy dolárov v roku 2020 v prípade Facebooku⁵⁰, 3,7 miliardy dolárov v prípade Twitteru, z toho z reklamy 3,2 miliardy⁵¹ a 1,9 miliardy v prípade TikToku⁵²) pochádza z **predaja online reklamy**. Čím viac času trávi užívateľ na online platforme, tým viac reklamy mu platforma potenciálne môže ukázať.

Veľmi veľké online platformy sa tak uchádzajú o **pozornosť** svojich užívateľov, čo je obmedzený zdroj. Pozornosť je podľa Encyklopédie Britannicy v psychológii definovaná ako „koncentrácia vnímania na nejaký fenomén vylúčením iných stimulov“.⁵³ Vzhľadom na zvyšujúci sa význam pozornosti v ekonomických vzťahoch v spoločnosti hovorí Goldhaber a po ňom ďalší o **ekonomike pozornosti** (angl. *attention economy*)⁵⁴ - hoci sa naša doba niekedy označuje ako doba informačnej ekonomiky, podľa Goldhabera je to nesprávne, keďže v ekonómii (a ekonomike) ide o to, ako spoločnosť rozdeľuje obmedzené zdroje, čo informácie nie sú, zatiaľ čo pozornosť áno.

Online platformy sa teda snažia získať čo najviac z tohto obmedzeného zdroja - pozornosti svojich užívateľov - a premeniť ho na svoj príjem v podobe predaja online reklamy.

Využívajú pritom tri prístupy:

1. Poskytnutie **atraktívnych užívateľských rozhraní s prvkami návykovosti** využívajúcimi **persuazívne techniky** založené na poznatkoch z behaviorálnej psychológie, ktoré zahŕňajú napr. tzv. *push* notifikácie (o pridaní nového príspevku priateľom, o akceptácii priateľstva atď.), informáciu o prečítaní správy príjemcom, nekonečné skrolovanie a pod.⁵⁵
2. Poskytnutie **zaujímavého, pútavého obsahu** - na online platformách je viac obsahu, ako je možné v obmedzenom (konečnom) čase skonzumovať. Je preto potrebné zobrazené informácie filtrovať, na čo sa využíva **personalizované odporúčanie**, t. j. informácie (príspevky na online platforme) sú zoradené podľa ich odhadovanej

⁵⁰ Facebook Revenue and Usage Statistics (2021). Dostupné na: <https://www.businessofapps.com/data/facebook-statistics/>.

⁵¹ Twitter Revenue and Usage Statistics (2021). Dostupné na: <https://www.businessofapps.com/data/twitter-statistics/>.

⁵² TikTok Revenue and Usage Statistics (2021). Dostupné na: <https://www.businessofapps.com/data/tik-tok-statistics/>.

⁵³ MCCALLUM, W.C. "Attention". Encyclopedia Britannica, 9 Jun. 2015, Dostupné na:

<https://www.britannica.com/science/attention> V anglickom origináli definované ako "the concentration of awareness on some phenomenon to the exclusion of other stimuli."

⁵⁴ GOLDBERGER, M.H. Attention Shoppers! 1997, WIRED. Dostupné na: <https://www.wired.com/1997/12/es-attention/>.

⁵⁵ Pozri napr. článok o Tristanovi Harrisovi, bývalom zamestnancovi spoločnosti Google zodpovednom za filozofiu produktov BOSKER, B. The Binge Breaker. The Atlantic, November 2016. Dostupné na: <https://www.theatlantic.com/magazine/archive/2016/11/the-binge-breaker/501122/>.

relevancie / pútavosti pre konkrétného užívateľa. Online platforma vyberá (v prípade sociálnych médií) príspevky od užívateľov / z kanálov, ktoré užívateľ sleduje (ktoré odoberá, páčia sa mu a pod.), ale aj príspevky od iných užívateľov / z iných kanálov, o ktorých si myslí, že by daného užívateľa mohli zaujať. Vplyv na poradie príspevkov má aj ich sponzorovanie, resp. zaradenie platenej reklamy.

3. Poskytnutie **relevantných reklám** - cieľom online platformy nie sú len *impresie*, t. j. zobrazenie reklamy, ale keďže pozornosť užívateľa je obmedzená, snaží sa vybrať také, u ktorých je vyššia pravdepodobnosť, že na ne užívateľ zareaguje. Aj pri online reklame preto platformy využívajú metódy **personalizovaného odporúčania**.

Už niekoľko rokov prebieha diskusia o škodlivosti mechanizmov, na ktorých stojí ekonomika pozornosti, napr. na schopnosť ľudí sústrediť sa, vykonávať hlbokú prácu⁵⁶, ale aj na psychiku detí a dospelujúcich.⁵⁷ Napriek tomu nevyzerá, že by v tomto smere malo v najbližšom čase prísť k významnejšej zmene, aj keď niektoré platformy (napr. Spotify či YouTube Music) umožňujú užívateľom rozhodnúť sa, či budú za ich službu "platiť" svojou pozornosťou v podobe pozerania / počúvania reklám alebo si zaplatia prémiové predplatné.

Paralelne prebieha diskusia o **rizikách online platforiem v súvislosti s rozsiahlym zberom údajov o užívateľoch**, ktoré realizujú za účelom personalizovania odporúčaní. Nejde len o zaznamenávanie, na aké príspevky užívateľ klikol alebo interagoval v podobe zdieľania, komentovania a pod., ale často aj ako dlho užívateľ pozeral na konkrétny príspevok (čomu je prispôsobené aj rozhranie, aby bol viditeľný práve jeden príspevok pre jednoduchšie meranie), aké iné stránky užívateľ na webe navštívil (pomocou tzv. cookies a pluginov do webových stránok, napr. tlačidlo "Páči sa mi to"), údaje zo zariadenia, z ktorého užívateľ na platformu pristupuje a pod.

Takýto zber je podľa Shoshany Zuboff základom tzv. **kapitalizmu dohľadu** (angl. *surveillance capitalism*), v ktorom sú tradičné vzťahy industriálneho kapitalizmu nahradzované alebo dopĺňané novými praktikami založenými na neobmedzenej „ťažbe“ behaviorálnych údajov a následnej monetizácii predpovedania správania užívateľov a manipulácii ich správania.⁵⁸

Ekonomika pozornosti môže fungovať aj mimo online platforiem (napokon, keď o nej v roku 1997 Goldhaber písal článok, veľmi veľké platformy a sociálne médiá, ako poznáme dnes, ešte neexistovali), ale zdá sa, že tieto významne zefektívňujú rozdeľovanie obmedzeného zdroja, ktorým je pozornosť užívateľov. Naopak, kapitalizmus dohľadu, o ktorom hovorí Zuboff, predpokladá sústredenie veľkého množstva údajov o veľkom množstve užívateľov v rukách pár veľkých spoločností, ktoré majú miestami až monopolné postavenie na trhu. Neponúka

⁵⁶ Tamže.

⁵⁷ Podľa nedávno uniknutých interných dokumentov zo spoločnosti Facebook (tzv. *Facebook Files*) spoločnosť vie už dlhšiu dobu vie o negatívnom vplyve Instagramu na psychické zdravie dospievajúcich dievčat, napr. v podobe zvýšeného rizika porúch príjmu potravy, depresii a pod. WELLS, G. - HORWITZ, J. - SEETHARAMAN, D. Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show. *The Wall Street Journal*, 2021. Dostupné na: <https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739>.

⁵⁸ ZUBOFF, S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs, 1. vydanie, 2019.

východiská, čo s ekonomikou pozornosti, ale zameriava sa na kroky smerujúce k zníženiu vplyvu veľmi veľkých online platforiem, napr. v podobe regulácií.

Riziká vyplývajúce zo zberu veľkého množstva osobných údajov sú už pomerne známe a v rámci EÚ aj regulované (v podobe GDPR). Metódy odporúčania, ktoré online platformy využívajú pre získanie pozornosti užívateľov (a kvôli ktorým prístupujú k zberu údajov o užívateľoch), ako aj benefity a možné riziká z nich vyplývajúce, sú však mimo odbornej komunity známe oveľa menej.

2.2 Metódy odporúčania

Odporúčacie systémy vznikli ako reakcia na informačné zahltenie užívateľov najmä vo firemnom prostredí koncom 80. rokov minulého storočia s cieľom redukovať množstvo informácií. Postupným vývojom technológií tieto snahy prechádzali cez 3 vývojové etapy.⁵⁹

V prvej vývojovej etape užívateľ dostal len časť prichádzajúcich informácií: v analógii z firemného prostredia by mu bola presmerovaná časť prichádzajúcich emailov, napr. 20%. Aj keď sa týmto spôsobom zníži množstvo informácií, stále sa tu nachádza množstvo nerelevantných pre daného užívateľa. V ďalšej etape vývoja sa nasadil filter, ktorý prefiltroval výstup z prvej etapy o kľúčové slová a vyradil časť nerelevantných informácií (napr. spam alebo komunikáciu nerelevantnú k biznisu firmy). Poslednú etapu vývoja predstavujú samotné personalizované odporúčače, ktoré z celej množiny informácií odporúčajú len tie relevantné pre konkrétneho užívateľa. Opäť v analógii firemného prostredia: užívateľ dostane len emaily, ktoré sú relevantné pre jeho pozíciu a neho ako osobu.

Vo všeobecnosti odporúčacie systémy definujeme ako softvérové nástroje, ktoré generujú návrhy položiek pre konkrétneho užívateľa.⁶⁰

S rozvojom užívateľmi generovaného obsahu (napr. sociálne siete, blogy) sa odporúčacie systémy plynule rozšírili aj do iných ako firemných aplikácií - najmä online služieb - a dnes sú (odporúčacie systémy) prakticky zlatým štandardom pri užívateľskej interakcii s webovými aplikáciami. S postupným rozmachom sa rozširovali aj ciele, s ktorým odporúčacie systémy boli nasadzované: z primárnej funkcie redukovania informačného zahltenia užívateľov, cez zvyšovanie lojality zákazníkov, ponúkanie rôznorodého obsahu, až po zvyšovanie predaja produktov.

Bez ohľadu na typ odporúčania a metódu, ktorá je pre generovanie návrhov položiek využitá, platí nutná podmienka, ktorá musí byť splnená - odporúčacia metóda musí mať k dispozícii informáciu o preferenciách užívateľov. Pri interakcii so systémom, resp. webovou aplikáciou, užívateľ zanecháva stopy (tzv. spätnú väzbu) v rôznych podobách (napr. videnia, hodnotenia,

⁵⁹ GOLDBERG, D. - NICHOLS, D. - OKI, B.M. - TERRY, D. Using collaborative filtering to weave an information tapestry. Commun. ACM 35, 12, December 1992, s. 61–70. Dostupné na: <https://doi.org/10.1145/138859.138867>.

⁶⁰ RICCI, F. - ROKACH, L. - SHAPIRA, B. Introduction to recommender systems handbook. In Recommender systems handbook, Springer, 2011, s. 1–35. Dostupné na: https://doi.org/10.1007/978-0-387-85820-3_1.

komentáre, nákupy). Práve spätná väzba sa používa ako indikátor preferencie užívateľov (napr. hodnotenie produktu).

2.2.1 Personalizácia a adaptácia

Pri nazeraní na metódy odporúčania rozlišujeme dva základné koncepty:

- *Personalizované odporúčanie* - návrhy sú vygenerované priamo pre konkrétneho užívateľa na základe jeho predchádzajúceho správania (resp. preferencií).
- *Nepersonalizované odporúčanie* - je všeobecne platné odporúčanie pre užívateľov služby alebo daného produktu (napr. odporúčaný spôsob prípravy jedla, odporúčaná postupnosť vzdelávacích materiálov, najčítanejšie správy).

Nepersonalizované odporúčania sa nachádzajú na prakticky všetkých webových sídlach, pričom si to užívatelia málokedy uvedomujú. Napríklad zoznam najčastejšie navštevovaných stránok, prípadne najčítanejšie články sú formou nepersonalizovaného odporúčania.

Na pomedzí medzi personalizovaným a nepersonalizovaným odporúčaním stojí *adaptácia* - *prispôsobenie* sa konkrétnej skupine užívateľov, napríklad zobrazenie tutoriálu novým užívateľom alebo upozornenie na nový film fanúšikom preferujúcim daný žáner. Inými slovami, pojem adaptácia priamo smeruje na prispôbovanie obsahu, resp. rozhrania. Zatiaľ čo pri odporúčaní užívateľovi prezentujeme množinu návrhov - odporúčaní, v prípade adaptácie dochádza priamo k úprave obsahu, resp. rozhrania, s ktorým užívateľ interaguje. Napríklad dôjde k preusporiadaniu položiek navigácie (menu).

Z pohľadu tejto štúdie sa budeme bližšie venovať personalizovaným odporúčaniam, vzhľadom na ich najväčší dopad na jednotlivcov, možnosti a rozšírenie v online platformách.

2.2.2 Modelovanie užívateľa

Nutnou podmienkou pre generovanie personalizovaného odporúčania je znalosť preferencií užívateľov (resp. predchádzajúcich akcií užívateľa). Na účel uchovávaní a modelovania preferencií užívateľov v online platformách sa používajú modely užívateľov. V základnej podobe je model užívateľa tvorený dvojicou [klúč, hodnota]⁶¹, kde klúč predstavuje danú preferenciu (napr. kategóriu správ - šport) a hodnota kvantifikuje preferenciu užívateľa. Inými slovami, model užívateľa charakterizuje daného užívateľa v špecifickom čase na základe informácií pozorovaných v určitom systéme (webovej aplikácii).

Napriek mnohým obmedzeniam je v online platformách zaužívaný behaviorálny prístup, kedy predpokladáme, že správanie užívateľa je podmienené jeho preferenciami. Na základe tohto predpokladu je modelovanie užívateľa založené na pozorovaní jeho správania - akcií, ktoré v systéme vykonal (resp. nevykonal). Jedná sa o zjednodušenie, ktoré umožňuje generovať odporúčania aj bez reálnej znalosti preferencií daného užívateľa.

⁶¹ SENOT, CH. - KOSTADINOV, D. - BOUZID, M. - PICAULT, J. - AGHASARZAN, A. - BERNIER, C. Analysis of strategies for building group profiles. User Modeling, Adaptation, and Personalization, Springer Berlin Heidelberg., 2010, s. 40–51. Dostupné na: https://doi.org/10.1007/978-3-642-13470-8_6.

Rozlišujeme dva základné zdroje informácií:

- *Implicitná spätná väzba* - odvodené informácie z akcií užívateľa. Napríklad na základe počtu prečítaných článkov o športe odvodíme jeho preferenciu pre danú kategóriu.
- *Explicitná spätná väzba* - priame vyjadrenie hodnotenia na konkrétny prvok, resp. preferenciu. Užívateľ hodnotí daný prvok (napr. ohodnotí, že sa mu práve prečítaný článok o športe páči, resp. je preňho relevantný), prípadne explicitne označí, že si želá dostávať články o športe.

V závislosti od zložitosti použitého modelu užívateľa rastie množstvo a typ informácií reprezentovaných v modeli:

- Demografické informácie
- Ciele a úlohy
- Záujmy
- Zručnosti a kompetencie
- Charakteristiky - osobnostné, kognitívne a pod.
- Nálada

Vzhľadom na rôzne domény a spôsoby aplikácie môže jeden systém využívať viacero modelov užívateľa, ktoré zohľadňujú aj časový aspekt preferencií užívateľa - napr. krátkodobé a dlhodobé. Bez ohľadu na typ modelu užívateľa je modelovanie kontinuálny proces, ktorý sa skladá z troch základných krokov:

- *Zber dát* - hodnotenia, záznamy o interakcii, kontext,
- *Inferencia (usudzovanie)* - odvodenie znalostí, preferencií, cieľov a pod. na základe zozbieraných informácií,
- *Adaptácia, resp. personalizácia* - aplikovanie metódy odporúčania na základe modelu užívateľa.

V jednotlivých aplikačných doménach sa význam modelu užívateľa líši. Napríklad vo výučbových systémoch je nevyhnutné modelovať znalosti konkrétnych študentov. V prípade elektronického obchodu si základné odporúčače vystačia s informáciou o predchádzajúcich nákupoch. Vo všeobecnosti však platí, čím komplexnejší model užívateľa, tým sú generované odporúčania kvalitatívne lepšie.

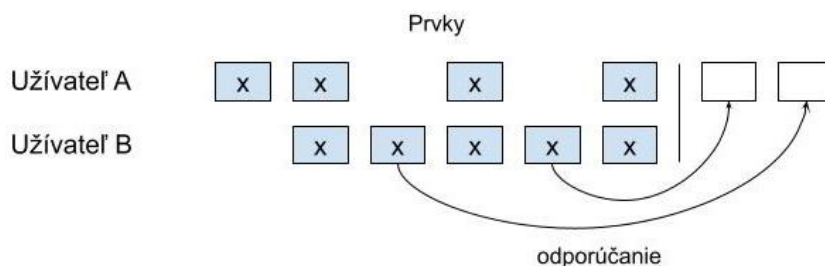
2.2.3 Personalizované odporúčanie

Prístupy personalizovaného odporúčania vo všeobecnosti využívajú metódy strojového učenia a štatistické prístupy. Aktuálne smery výskumu viac rozvíjajú metódy strojového učenia, najmä neurónové siete.

Základné delenie prístupov personalizovaného odporúčania stavia na analógii z bežného života a správania ľudí, ktorí majú tendenciu zvažovať rady svojho okolia:

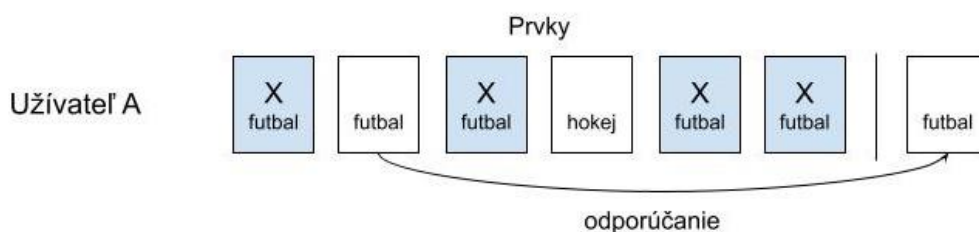
- *Kolaboratívne filtrovanie* - generuje odporúčania na základe podobnosti užívateľov. Vychádza z predpokladu, že ak sa preferencie (resp. história) dvoch užívateľov zhodovali

alebo podobali v minulosti, je vysoko pravdepodobné, že sa ich preferencie budú zhodovať alebo podobať aj v budúcnosti (obrázok č. 1). Podstatu kolaboratívneho filtrovania tvoria informácie o preferenciách užívateľov. Aby sme dosiahli uspokojujúce výsledky, je nevyhnuté, aby systém používalo veľké množstvo užívateľov, čo zvyšuje šancu, že nájdeme užívateľov s podobnými preferenciami. Výhodou kolaboratívneho filtrovania je abstrakcia od obsahu odporúčaných prvkov, o ktorom nemusíme nič vedieť (pre fungovanie metód stačí jednoznačná identifikácia prvkov).



Obrázok č. 1
Princíp fungovania kolaboratívneho filtrovania
Zdroj: Vlastná tvorba autorov.

- **Obsahové odporúčanie** - generuje odporúčania na základe analýzy obsahu prvkov. Odporúča prvky podobné tým, ktoré sa užívateľovi páčili v minulosti (obrázok č. 2). Podobnosť medzi prvkami bola historicky definovaná ako textová podobnosť (napr. spravodajských článkov), prípadne textová podobnosť metadát prvkov (režisér, žáner a pod.). V dnešnej dobe už nie sú kladené špeciálne požiadavky na obsah vzhľadom na posun v spracovaní obrazu či zvuku. Obsahové odporúčanie je tak možné aplikovať do nových domén, pričom pre jeho generovanie stačí poznať preferencie daného užívateľa a spracovať dostupný obsah. V porovnaní s kolaboratívnym odporúčaním nie je teda potrebné analyzovať správanie ostatných užívateľov.



Obrázok č. 2
Princíp fungovania obsahového odporúčania
Zdroj: Vlastná tvorba autorov.

Na druhej strane je zrejmé, že obsahové odporúčanie zo svojej podstaty konverguje k uzavretiu užívateľa do filtračnej bubliny a následne degraduje. Je preto nevyhnutné implementovanie mechanizmov, ktoré zabezpečia diverzifikáciu odporúčaných prvkov. Inou možnosťou je kombinácia viacerých typov odporúčania na danej online platforme (čo predstavuje zlatý štandard moderných webových aplikácií).

- *Hybridné prístupy* - vychádzajú z predpokladu, že kombináciou viacerých metód (nie nutne iných prístupov) získame "lepšie" výstupy. Zlepšenia pritom môžu smerovať na rôzne aspekty (napr. presnejšie odporúčania, vyššiu diverzitu). Hybridný prístup spravidla smeruje k redukcii ohraničení a problémov partikulárnych metód, ktoré zahŕňa. Existuje pomerne veľké množstvo spojení dvoch a viacerých metód. Medzi najjednoduchšie patria prepínacie hybridné systémy, ktoré na základe explicitného pravidla (napr. množstvo hodnotených prvkov) prepínajú medzi viacerými metódami odporúčania. Iným príkladom je kaskádový hybridný systém, kde prvá metóda vygeneruje kandidátov na odporúčanie a druhá tento zoznam preusporiada a upraví.

Okrem vyššie opísaného základného delenia rozoznávame ďalšie dimenzie, na základe ktorých rozlišujeme aj iné typy odporúčačov. Ak odporúčací systém uvažuje aj kontextuálne informácie (napr. polohu užívateľa, zariadenie, z ktorého pristupuje, počasie, ročné obdobie), hovoríme o *kontextovo-založenom odporúčaní*.

So zavedením GDPR (a nárastom užívateľov, o ktorých nemáme dostupné informácie - ich preferencie) sa čoraz viac stáva populárnym *odporúčanie v rámci sedení*. Sedením chápeme postupnosť akcií, ktorú užívateľ vykoná v rámci jednej návštevy nejakého webového sídla či služby. Pri tomto prístupe metódy uvažujú najmä aktuálne akcie užívateľa bez zohľadnenia historických preferencií. Tým pádom je možné generovať odporúčania aj neidentifikovaným užívateľom, o ktorých nemáme predchádzajúce informácie.

2.2.4 Vyhodnocovanie úspešnosti

Pri vyhodnocovaní úspešnosti metód personalizovaného odporúčania sa typicky používajú metriky prevzaté z oblasti vyhľadávania informácií a strojového učenia. Je dôležité poznamenať, že tieto metriky adresujú kvalitu generovaných odporúčaní z technického hľadiska, pričom ignorujú iné dôležité aspekty, ktoré hovoria o širšej kvalite, resp. adresujú biznisové otázky spojené s nasadením odporúčačov.

Základné metriky pre charakteristiku "úspešnosti" odporúčania sú:

- *Presnosť* - je charakterizovaná ako pomer správne odporúčaných prvkov voči všetkým odporúčaným prvkom. Napríklad, ak odporúčame zoznam 10 prvkov, z ktorých užívateľ označí za relevantné 2, dosiahne toto odporúčanie presnosť $2/10 = 20\%$.
- *Pokrytie* - je doplnková metrika k metrike presnosť. Vyjadruje pomer relevantných odporúčaní voči všetkým relevantným prvkom pre daného užívateľa. Ak teda vieme, že na online platforme sa nachádza 5 relevantných prvkov pre daného užívateľa, a užívateľ z odporúčaní identifikuje 2 z nich, pokrytie bude predstavovať $2/5 = 40\%$.
- *Správnosť* - reflektuje správne odporúčané (relevantné) prvky a správne neodporúčané (nerelevantné) prvky. Vzhľadom na veľký nepomer medzi počtom relevantných a nerelevantných prvkov v reálnych systémoch (relevantných je rádovo menej ako nerelevantných) metrika reportuje vysoké hodnoty aj pri kvalitatívne zlých odporúčaníach. Napríklad majme na online platforme celkovo 100 prvkov, pričom len 5 je relevantných pre užívateľa. Vygenerujeme 10 odporúčaní, ktoré budú zahŕňať 1 správne odporúčaný (relevantný) prvok a 9 nesprávne odporúčaných (nerelevantných). Zo zostávajúcich 90

prvkov sme správne neodporučili 86 nerelevantných prvkov a nesprávne 4 relevantné prvky. V tomto prípade dosiahneme správnosť $(1+86)/100 = 87\%$, a to napriek tomu, že 90% odporúčaných prvkov bolo nesprávnych (presnosť 10%) a odporúčaním sme pokryli len $1/5 = 20\%$ zo všetkých relevantných prvkov na platforme.

Keďže vyššie spomínané metriky ignorujú usporiadanie prvkov (nezáleží na ich poradí), čo nereflektuje zaužívané správanie užívateľov (užívatelia štandardne prechádzajú zoznamy smerom zhora nadol, a predpokladajú, že relevantné prvky budú na vrchu usporiadaného zoznamu), ďalšia skupina metrík zohľadňuje aj pozície odporúčaní:

- *Normalizovaný znížený kumulatívny zisk (nDCG)* - zohľadňuje usporiadanie prvkov pri odporúčaní a porovnáva ich s ideálnym usporiadaním. Ideálne usporiadanie predstavuje zoznam, v ktorom je najrelevantnejší prvok na prvom mieste nasledovaný druhým najrelevantnejším atď.
- *Priemerná presnosť (aP)* - sa používa na porovnávanie dvoch usporiadaných zoznamov odporúčaní, pričom za lepšie odporúčanie považujeme také, ktoré vygenerovalo relevantné prvky na prvé miesta zoznamu (obrázok č. 3).

Pozícia	Relevancia	Pozícia	Relevancia
1		1	x
2	x	2	
3	x	3	
4	x	4	x
5		5	x
aP = 0.64		aP = 0.7	

Obrázok č. 3

Porovnanie dvoch usporiadaných zoznamov personalizovaného odporúčania prostredníctvom metriky priemernej presnosti – mierne vyššie hodnotenie aP spôsobila relevantná prvá pozícia (vpravo)

Zdroj: Vlastná tvorba autorov.

Tretou skupinou metrík sú metriky vyjadrujúce odchýlku od hodnotenia užívateľa. V prípade, že užívateľ hodnotí odporúčaný prvok (napr. počet hviezdíčiek), tieto metriky vyjadrujú chybu odhadu odporúčača:

- *Priemerná absolútna chyba* - je vyjadrená ako priemerný rozdiel medzi predpoveďou a reálnym hodnotením užívateľa. Je zrejmé, že chyba je definovaná len pre prvky, ktoré užívateľ ohodnotil, preto nedáva úplný obraz o úspešnosti systému.
- *Priemerná kvadratická chyba* - v porovnaní s priemernou absolútnou chybou zohľadňuje viac väčšie chyby v odhade (vďaka ich umocneniu).

Vo všeobecnosti sa na vyhodnotenie kvality metódy odporúčania nepoužíva len jedna metrika, ale skupina metrík, ktoré dokážu systém opísať z viacerých hľadísk. Kvalitu odporúčaní ako takých vieme opísať aj z pohľadu ďalších "mäkkých" metrík:

- *Dôvera* - miera dôvery užívateľa v odporúčaciu metódu. Pri generovaní relevantných odporúčaní silnie miera dôvery v odporúčač ako taký.
- *Novosť* - predstavuje odporúčania, ktoré sú pre daného užívateľa nové. Napríklad, je málo pravdepodobné, že užívateľ vyhodnotí aktuálne masovo promovány filmový trhák ako nový.
- *Diverzita* - vyjadruje mieru rôznosti generovaných odporúčaní. Metrika pomáha odhaľovať filtračné bubliny, kedy odporúčač generuje tematicky identické odporúčania.
- *Chvost (Tail)* - opisuje schopnosť systému odporúčať prvky z takzvaného "dlhého chvosta". V systéme sa nachádza malé množstvo veľmi žiadaných prvkov a veľké množstvo prvkov, ktoré vykazujú nízku mieru interakcie (napr. užívatelia nevedia, že také prvky existujú). Schopnosť odporúčať tieto prvky so sebou často prináša ekonomické výhody a zároveň prispieva k vyššej diverzite.

V závislosti od dôvodu nasadenia odporúčacieho systému môžeme sledovať aj jeho nepriame efekty na užívateľov (napr. na spokojnosť užívateľov s online platformou). Z dlhodobého hľadiska tak môžeme sledovať mieru straty zákazníkov.

Medzi typické biznisové metriky, ktoré priamo vyjadrujú aj ekonomickú stránku odporúčacej metódy, patria:

- *Obrat* - je zvyčajne sledovaný v elektronických obchodoch. Odporúčací systém môže byť nasadený s cieľom reflektovať vybraný problém - napríklad s cieľom zaujať nových užívateľov. Zo strednodobého a dlhodobého hľadiska tak vieme vyhodnotiť príspevok odporúčania k zvýšeniu obratu.
- *Podiel klikanosti (angl. click-through rate, CTR)* - predstavuje štandardnú metriku v online platformách nezávisle na doméne. Predstavuje pomer "kliknutí" a zobrazení - v našom prípade odporúčaní daného prvku. Staví na predpoklade, že klik užívateľa predstavuje implicitne pozitívnu spätnú väzbu.
- *Strávený čas (time spent)* - umožňuje vyhodnotiť záujem užívateľa na základe času, ktorý strávi na danej platforme, prípadne konkrétnom prvku. Napríklad v doméne spravodajstva môže zohľadnenie metriky času odfiltrovať články, ktoré užívateľ navštívil, ale nečítal, od takých, ktoré skutočne čítal a zaujali ho. Pri pohľade na celú platformu môže striktná optimalizácia na čas strávený užívateľmi vyvolať nežiaduce správanie modelu - jeho optimalizáciu iba na túto metriku.

Zvolená metrika resp. skupina metrík umožňujú redukovať možné nežiaduce efekty zapríčinené strojovým prístupom založenom na spracovaní veľkého množstva dát. Sledovaním viacerých kritérií je možné predchádzať generovaniu nezaujímavých, úzko špecializovaných, prípadne nedôveryhodných odporúčaní.

2.2.5 Typické problémy odporúčacích systémov

Z pohľadu odporúčacích metód sa stretávame s typickými problémami, ktoré priamo vychádzajú z podstaty odporúčania ako takého, resp. zo samotnej idey daného typu odporúčacej metódy.

Studený štart je situácia, kedy metóda nedokáže vygenerovať odporúčania užívateľovi, ktorého preferencie nepoznáme (nový užívateľ). Je zrejmé, že rovnaký problém nastáva aj pri novom prvku (nový článok, produkt), ktorý ešte nezískal hodnotenia a teda systém ho nevie odporúčať. Rôzne metódy reagujú na nových užívateľov/prvky rôzne citlivo. Zatiaľ čo pri obsahovom odporúčaní je možné generovať odporúčania už po prvej interakcii užívateľa, pri kolaboratívnom filtrovaní pretrváva problém studeného štartu dlhšie.

Čierne a šedé ovce predstavujú užívateľov problematických z pohľadu metódy odporúčania. Za čierne ovce sú označovaní užívatelia, ktorí sa správajú (resp. ich preferencie sú) odlišne od ostatných užívateľov, a preto im nie je možné odporúčať kolaboratívnymi prístupmi. Za šedé ovce označujeme užívateľov, ktorých preferencie sa neustále menia, resp. nekorelujú so žiadnou zo skupín užívateľov. Pre takýchto užívateľov je opäť problematické generovať odporúčania kolaboratívnymi prístupmi.

Úzka špecializácia je stav odporúčacieho modelu, ktorý je úzko fokusovaný na konkrétnu preferenciu užívateľa. Napríklad mu ponúka len spravodajské články z jednej čiastkovej domény a ignoruje jeho iné preferencie. Užívateľ sa tak dostáva do informačnej bubliny, z ktorej sa len interakciou s odporúčacím systémom nevie dostať.

Riedkosť dát je typický problém odporúčacích metód. Vzhľadom na aplikačné domény sa odporúčacie systémy nasadzujú v systémoch s veľa užívateľmi a veľa prvkami, ktoré je potrebné odporúčať. Z povahy správania užívateľov však systém pozná (resp. dokáže odhadnúť) preferenciu užívateľa len pre veľmi obmedzenú množinu prvkov. Napríklad v doméne filmov existujú tisíce filmov, avšak jeden užívateľ štandardne hodnotí maximálne desiatky z nich. Vo výsledku tak riedkosť dát bežne dosahuje 99+%.

2.2.6 Príklady použitia na online platformách

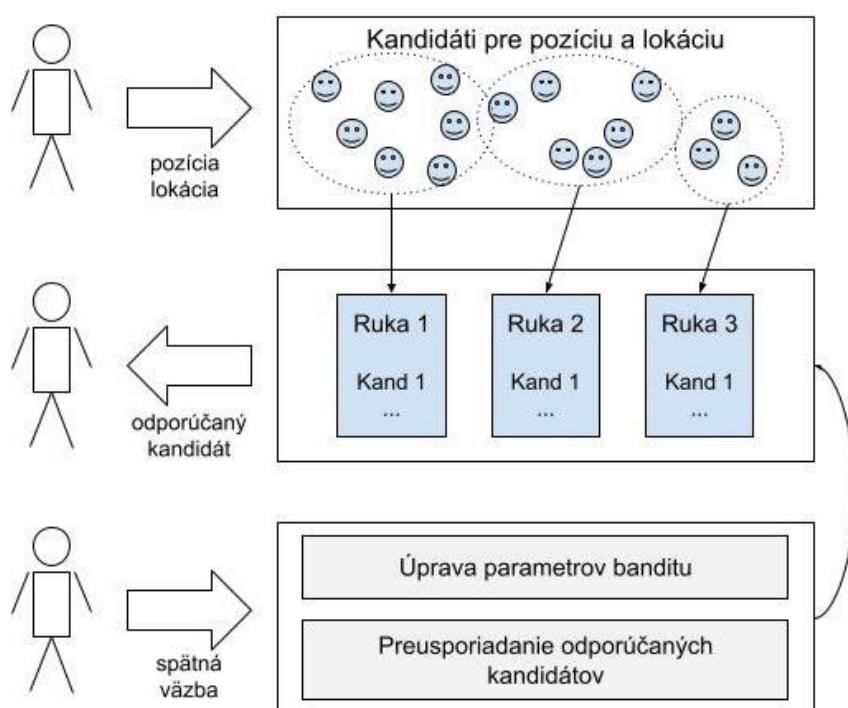
Odporúčacie systémy sú neoddeliteľnou súčasťou moderných webových aplikácií. Vzhľadom na množstvo užívateľmi generovaného obsahu umožňujú redukovať informačného zahltenia a v závislosti od aplikácie prinášajú príležitosti pre poskytovateľov a užívateľov zároveň.

Napríklad profesná sieť LinkedIn využíva personalizované odporúčania na základe profilov kandidátov prakticky pri všetkých produktoch, ktoré poskytuje na svojej online platforme - odporúčanie kontaktov, kanálov na odoberanie, spoločností a pod. Pre doménu pracovných ponúk sa však odporúčania negenerujú jednostranne, ale ide o špeciálny prípad obojstranného odporúčania. Podobne ako pri odporúčaní na zoznamovacích platformách, nestačí, že kandidát je vhodný pre ponuku, musí aj daná ponuka zaujať konkrétno kandidáta.⁶² Model siete LinkedIn využíva dvojstupňovú architektúru, kedy sa v prvom kroku usporiadajú kandidáti vyhovujúci danej ponuke podľa relevancie, v druhom kroku sa preusporiadajú *top-N* relevantní kandidáti na základe aktuálnych dynamických preferencií s využitím metódy tzv. mnohorukých banditov.

⁶² AI Behind LinkedIn Recruiter Search and Recommendation Systems. Dostupné na: <https://engineering.linkedin.com/blog/2019/04/ai-behind-linkedin-recruiter-search-and-recommendation-systems>.

Architektúra systému využívajúceho metódu mnohorukých banditov je znázornená na obrázku č. 4. Metódy mnohorukých banditov môžeme jednoducho priblížiť na príklade problému z reálneho života. Predstavme si hráča, ktorý stojí pred K hracími automatmi (tzv. jednorukými banditmi). Hráč stojí pred otázkou, ktorý z hracích automatov použije, pričom sa logicky snaží optimalizovať výhru. Hráč sa v priebehu hry musí rozhodovať, ako dlho bude skúšať jeden automat, kedy vyskúša iný a pod. Musí nájsť optimum medzi preskúmvaním iných možností a skúšaním konkrétneho automatu.

V reči online platforiem je problém analogický - máme množstvo obsahu (napr. príspevkov) a potrebujeme zistiť, ktorý prvok máme zobrazit' danému užívateľovi, o ktorom nemáme informácie, v akom poradí a čase. Jedná sa o komplexný optimalizačný problém, keďže systém dostáva informáciu len o "automatoch", ktoré si užívateľ vybral a nič o ostatných.



Obrázok č. 4

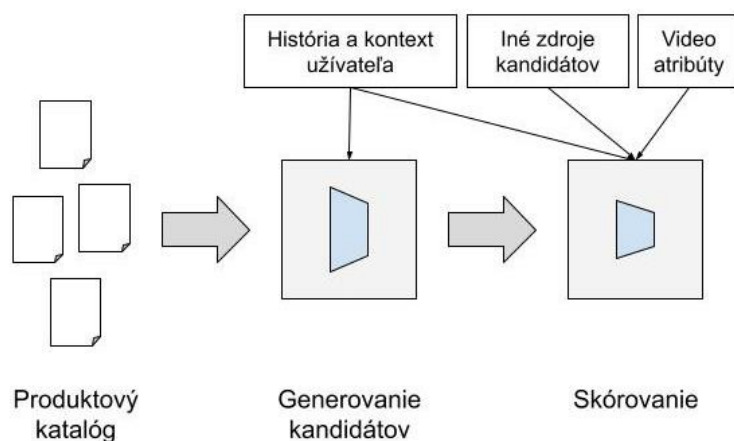
Architektúra personalizovaného odporúčania pomocou algoritmu mnohorukých banditov - LinkedIn.

Zdroj: Inšpirované.⁶³

V závislosti od veľkosti online platformy musí byť architektúra odporúčacieho systému navrhnutá s ohľadom na množstvo prvkov a užívateľov používajúcich systém. Napríklad platforma YouTube využíva dvojstupňový prístup (obrázok č. 5).⁶⁴

⁶³ Tamže.

⁶⁴ COVINGTON, P. - ADAMS, J. - SARGIN, E. Deep Neural Networks for YouTube Recommendations. In Proc. of the 10th ACM Conf. on Recommender Systems (RecSys '16). Association for Computing Machinery, New York, NY, USA, 2016, s. 191–198. Dostupné na: <https://doi.org/10.1145/2959100.2959190>.



Obrázok č. 5

Architektúra odporúčacieho modelu online platformy YouTube.

Zdroj: Inšpirované.⁶⁵

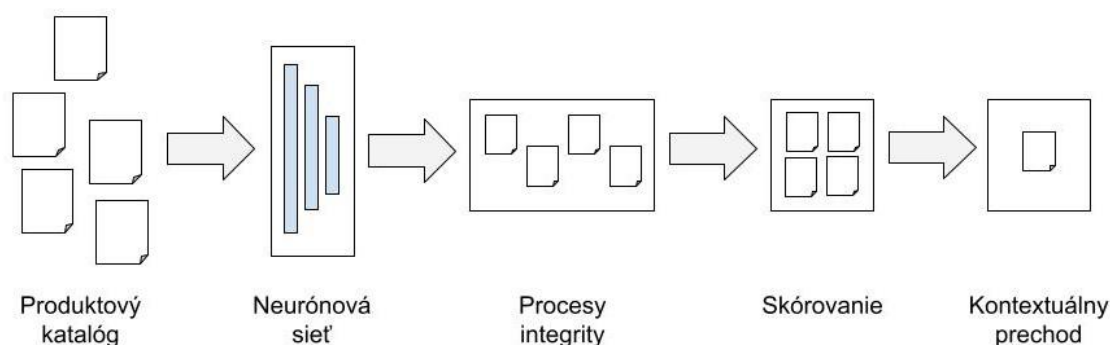
V prvom kroku sú z dostupných prvkov (videí) vygenerované stovky kandidátov na základe histórie daného užívateľa. V tomto kroku je využitý prístup kolaboratívneho filtrovania rozšíreného o zohľadnenie demografie užívateľov, resp. v súčasnosti pravdepodobne hlboké učenie. V druhom kroku je množina vygenerovaných kandidátov usporiadaná na základe ohodnocovacej funkcie, ktorá je definovaná platformou. V tomto kroku sa medzi vygenerovaných kandidátov môžu dostať aj prvky z iných zdrojov. Podobne sa v tomto kroku zohľadňujú aj ďalšie aspekty jednotlivých prvkov, napr. ich popularita a vývoj v čase, relevantnosť daného kanála a história interakcie s ním, jazyk a pod.

Podobnú dvojstupňovú architektúru využíva aj personalizovaný algoritmus radenia príspevkov na online platforme Facebook⁶⁶ (obrázok č. 6). V prvom kroku sú zozbierané všetky relevantné prvky pre daného užívateľa (na základne jeho sociálnej siete a odberov, ktoré má aktívne). Ako kandidáti sú vybrané nie len prvky, ktoré vznikli od posledného prihlásenia, ale aj prvky, ktoré daný užívateľ ešte nevidel. Okrem toho, prvky s ktorými užívateľ interagoval, ale od tejto interakcie boli aktívne (napr. vznikla k nim diskusia), sú tiež zaradené medzi kandidátov.

V ďalšom kroku je každý prvok oskórovaný na základe modelov hlbokého učenia, pričom sú jednotlivé charakteristiky lineárne kombinované (skóre z hodnotení, komentárov a pod). V prvom prechode sú aplikované pravidlá - napríklad na zabezpečenie rôznorodosti obsahu (redukovanie viacerých repetitívnych videí a pod.). Z takto upravenej množiny vyberie model približne 500 finálnych kandidátov, ktorí sú finálne usporiadaní v záverečnom prechode (so zohľadnením aktuálneho kontextu).

⁶⁵ Tamže.

⁶⁶ News Feed Ranking. Dostupné na: <https://engineering.fb.com/2021/01/26/ml-applications/news-feed-ranking/>.



Obrázok č. 6

Architektúra personalizovaného usporiadania príspevkov na online platforme Facebook.

Zdroj: Inšpirované.⁶⁷

2.3 Dopady použitia odporúčacích systémov

Z pohľadu dopadu použitej technológie na jednotlivca sú v kontexte odporúčacích systémov dôležité dva hlavné aspekty a ďalšie odvodené:

Transparentnosť generovaných odporúčaní. Metódy umelej inteligencie a špeciálne strojového učenia často nedovoľujú jednoznačne a zrozumiteľne vysvetliť, ako bolo dané odporúčanie vygenerované (resp. prečo je odporúčaný prvok A, a nie prvok B); ide o otvorený výskumný problém.⁶⁸ Generovanie vysvetliteľných odporúčaní je v záujme samotných prevádzkovateľov.

Použitie hlbokého učenia a neurónových sietí problém s transparentnosťou generovaných odporúčaní ešte komplikuje, keďže pri tomto prístupe nemáme explicitne reprezentované preferencie užívateľa či model užívateľa. V posledných rokoch je čoraz viac vedeckej pozornosti venovanej práve generovaniu ľudske pochopiteľných vysvetlení pre koncových užívateľov (bolo dokázané, že vysvetľovanie odporúčaní zvyšuje dôveru užívateľov a ich zážitok pri interakcii so systémom⁶⁹).

Funkcia užitočnosti optimalizovaná v procese tréningu modelu. Tak ako iné technológie, natrénovaný model odporúčania optimalizuje, resp. reflektuje kritériá nastavené v procese tréningu. Práve funkcia užitočnosti predstavuje rozdiel medzi odporúčaniami maximalizujúcimi spokojnosť užívateľa a odporúčaniami maximalizujúcimi čas strávený na danej online platforme. Vo viacerých reálnych prípadoch je však funkcia užitočnosti doplnená o explicitne definované pravidlá. Inými slovami, zoznam odporúčaní vygenerovaných funkciou užitočnosti optimalizujúcou potreby užívateľa je doplnený o odporúčania optimalizujúce iné kritériá (ktoré môžu viesť ku negatívnym efektom na užívateľa).

⁶⁷ Tamže.

⁶⁸ Pozri napr. ZHANG, Y. - CHEN, X. Explainable Recommendation: A Survey and New Perspectives, Foundations and Trends® in Information Retrieval, vol. 14, no. 1, 2020, s. 1-101. Dostupné na: <https://doi.org/10.1561/15000000066>.

⁶⁹ TINTAREV, N. - MASTHOFF J. A Survey of Explanations in Recommender Systems, 2007 IEEE 23rd International Conference on Data Engineering Workshop, 2007, s. 801-810. Dostupné na: <https://doi.org/10.1109/ICDEW.2007.4401070>.

Napr. platforma YouTube podľa článku z roku 2019 (aktuálny stav preto môže byť iný) využíva viacero funkcií užitočnosti založených na metrikách interakcie užívateľa s platformou (kliknutie na odporúčané video, strávený čas a pod.) a metrikách spokojnosti užívateľa.⁷⁰ Váhy metrík sú nastavované manuálne, aby bola (slovami autorov článku) “dosiahnutá najlepšia výkonnosť z pohľadu interakcií užívateľa a jeho spokojnosti.”⁷¹

Uniknuté dokumenty zo spoločnosti Facebook ukazujú, že spoločnosť v roku 2018 zmenila funkciu užitočnosti, aby zamedzila poklesu interakcií užívateľov na platforme a podporila tzv. zmysluplné sociálne interakcie (angl. Meaningful Social Interactions, MSI); výsledkom síce bol nárast interakcií na platforme, ale pocit spokojnosti užívateľov poklesol.⁷² V rámci týchto zmien sa tiež väčšia váha pridala interakciám užívateľov, v ktorých vyjadrovali svoje emócie (napr. hnev, smiech a pod.) oproti neutrálnemu “Páči sa mi to”.⁷³ Výsledkom oboch týchto zmien bolo, že sa začal viac odporúčať polarizujúci obsah, ktorý má väčšiu pravdepodobnosť, že ho užívatelia budú komentovať, zdieľať a vyjadrovať k nemu (často negatívne) emócie.

Oba tieto prípady ilustrujú dôležitosť funkcie užitočnosti a nastavenia váh na výstupné odporúčania; bez ich znalosti nie je možné hovoriť o skutočnej transparentnosti a možnosti kontroly prípadných negatívnych dopadov. Zároveň sa ukazuje, že zmena funkcie užitočnosti alebo jej váh môže mať nezamýšľané dopady na odporúčaný obsah a správanie užívateľov, ktoré sa len ťažko dajú odhaliť bez systematického auditovania, resp. testovania, ktoré aktuálne pri online platformách, zdá sa, absentuje.

Uzatváranie užívateľa do informačných bublín. V prípade, že návrh odporúčacieho systému (napr. typ metódy, funkcia užitočnosti) ignoruje obmedzenia technológií, môže viesť jeho využívanie k uzatváraniu užívateľov do *informačných bublín*.⁷⁴ Na jednej strane sa užívatelia aj v reálnom živote obklopujú ľuďmi s podobnými názormi či záujmami. Z tejto podstaty je zrejmé, že odporúčací systém bude na online platforme odporúčať príspevky od takto podobných užívateľov (ktorí sú napr. priatelia). Na strane druhej sú niektoré metódy náchylné na odporúčanie podobného obsahu už zo svojej podstaty (napr. obsahové odporúčanie).

V súvislosti s informačnými bublinami sa hovorí aj o *komorách ozvien* (angl. *echo chambers*), čo je stav intelektuálnej izolácie, v ktorom sú rovnaké myšlienky a názory opakované, vzájomne potvrdzované a zosilňované v relatívne homogénnych skupinách.⁷⁵ Tieto komory vznikajú prirodzeným správaním ľudí opísaným vyššie, ale prispievajú k tomu aj algoritmy napr. uzatváraním užívateľov do informačných bublín.

⁷⁰ ZHAO, Z. et al. Recommending what video to watch next: A multitask ranking system. RecSys 2019 - 13th ACM Conference on Recommender Systems, 2019, s. 43–51. Dostupné na: <https://doi.org/10.1145/3298689.3346997>.

⁷¹ Tamže.

⁷² HAGEY, K. - HORWITZ, J. Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead. The Wall Street Journal, 2021. Dostupné na: <https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215>.

⁷³ MERRILL, J.B. - OREMUS, W. Five points for anger, one for a 'like': How Facebook's formula fostered rage and misinformation. The Washington Post, 2021. Dostupné na: <https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/>.

⁷⁴ Používa sa aj pojem filtračná bublina. Pozri napr. PARISER, E. The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think, 2012.

⁷⁵ BESSI, A. Personality traits and echo chambers on Facebook. Computers in Human Behavior 65, 2016, s. 319–324. Dostupné na: <https://doi.org/10.1016/j.chb.2016.08.016>.

Existujú prístupy na znižovanie tendencií odporúčacích algoritmov uzatvárať užívateľov do informačných bublín, napr. diverzifikácia⁷⁶ alebo prezentovanie rôznych pohľadov v spravodajstve,⁷⁷ avšak aj tieto majú svoje limitácie (napr. identifikácia rôznych pohľadov a stanovísk si vyžaduje pokročilé metódy prirodzeného jazyka a ide stále o otvorený výskumný problém), a nie vždy sú v súlade s hlavným cieľom platforiem vyplývajúcim z ekonomiky pozornosti.

Skreslenia a férovosť v odporúčaní. Ako sme naznačili už v predchádzajúcom bode, v dátach, z ktorých sa odporúčania generujú, sú prítomné rôzne skreslenia (angl. *biases*), ktoré často vznikajú na základe predsudkov a stereotypov prítomných v spoločnosti. Keďže odporúčací systém vychádza z takýchto dát, je takmer isté, že bude reflektovať toto skreslenie aj v generovaných odporúčaní. Napríklad, ak je v dátach prítomné skreslenie, že ženy získavajú slabšie platenú prácu, je z pohľadu funkcie užitočnosti (vygenerovať úspešné odporúčanie) optimálne odporučiť nižšie platenú prácu (pretože také je pozorovanie z bežného života). Takýmto spôsobom tieto skreslenia a rôzne stereotypy nechtiac kódujeme do systémov na podporu rozhodovania.

Skreslenia pritom nie sú spôsobené len dátami a učiacim algoritmom, ktorý si ich z dát osvojí, ale aj rozhraním a interakciou užívateľov s ním (napr. pri prezentovaní odporúčaní hrá významnú rolu tzv. *pozičné skreslenie*, pri ktorom sú s vyššou pravdepodobnosťou užívateľmi kliknuté položky, ktoré sa nachádzajú v zozname vyššie, ale aj *skreslenie spôsobené výberom* užívateľov, ktorí častejšie hodnotia položky, ktoré sa im páčia, ako tie, o ktoré nemajú záujem); ide pritom o posilňujúci sa cyklus, kedy jedno skreslenie generuje a posilňuje ďalšie.⁷⁸

Skreslenia a zabezpečenie férovosti predstavujú problém metód umelej inteligencie vo všeobecnosti, nielen odporúčacích systémov, ktorý sa prejavuje v rôznych doménach, ako napr. posudzovanie uchádzačov o zamestnanie či o povýšenie, posudzovanie finančnej situácie za účelom poskytnutia pôžičky a pod., ale aj na sociálnych médiách a iných online platformách.

Zbieranie spätnej väzby a odvodzovanie preferencií užívateľa v súvislosti s ochranou súkromia. Aj keď je zbieranie spätnej väzby, resp. akcií užívateľa konceptuálne oddelené od metód personalizovaného odporúčania, je zrejmé, že pre generovanie odporúčaní nutne potrebujeme informácie o preferenciách užívateľa. Tak ako v iných aplikačných doménach, predstavujú takéto informácie potenciálne riziko zneužitia. Je preto nevyhnutné podniknúť kroky na jeho minimalizáciu, čo môže mať viacero podôb - od zverejnenia toho, čo sa o užívateľovi zaznamenáva, cez umožnenie nastavenia rozsahu zbieraných údajov, až po uchovávanie údajov a odvodených preferencií (modelu užívateľa) na strane užívateľa (čomu

⁷⁶ KUNAVAR, M. - POŽRL, T. Diversity in recommender systems – A survey, Knowledge-Based Systems, vol. 123, 2017, s.154-162. Dostupné na: <https://doi.org/10.1016/j.knosys.2017.02.009>.

⁷⁷ DRAWS, T. - TINTAREV, N. - GADIRAJU, U. Assessing Viewpoint Diversity in Search Results Using Ranking Fairness Metrics. SIGKDD Explor. Newsl. 23, 1, June 2021, s. 50–58. Dostupné na: <https://doi.org/10.1145/3468507.3468515>.

⁷⁸ BAEZA-YATES, R. Bias on the web. *Communications of the ACM*, 61(6), 2018, s. 54–61. Dostupné na: <https://doi.org/10.1145/3209581>.

sa výskumne venuje napr. oblasť distribuovaných modelov užívateľov⁷⁹ alebo prístupy diferencovateľného súkromia⁸⁰).

Na záver tejto časti môžeme konštatovať, že odporúčacie systémy sú užitočná technológia, ktorej cieľom je znížiť informačné zahltenie užívateľov a umožniť im efektívnejší prístup k pre nich relevantným informáciám. Na druhej strane so sebou, podobne ako iné nasadenia systémov AI, prinášajú viaceré riziká negatívnych dopadov na užívateľov, a to vo všeobecnosti, ako aj špecificky v kontexte šírenia dezinformácií. Odporúčacie systémy sú pritom vo veľkom využívané online platformami v rámci ich modelu fungovania založenom na ekonomike pozornosti, v ktorom sú užívatelia využívaní ako produkt na zvýšenie zisku (skrz predaj online reklamy). Zároveň, veľmi veľké online platformy (špecificky sociálne média) sú dnes významným zdrojom informácií a spravodajstva pre veľkú časť populácie, čo zvyšuje riziko možných negatívnych dopadov. Vzhľadom na ekonomickú silu online platforiem a ich vplyv na spoločnosť preto vidíme jednoznačnú potrebu úpravy fungovania online platforiem a nimi používaných algoritmov prostredníctvom regulácie.

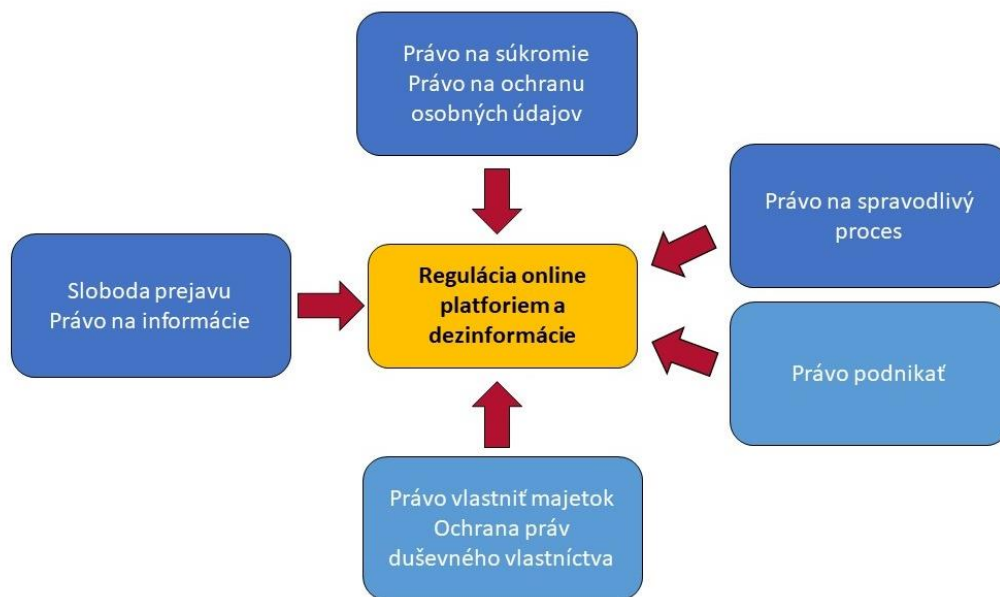
⁷⁹ Napr. GUO, H. - CHEN, J. - WU, W. - WANG, W. Personalization as a service: The architecture and a case study. In Proceedings of the 1st International Workshop on Cloud data management (CloudDB '09). ACM, New York, NY, USA, 2009, s. 1–8. Dostupné na: <http://pages.cs.wisc.edu/~wentaowu/papers/cikm09-clouddb-workshop.pdf>.

⁸⁰ Pozri napr.: CHEU A. - SMITH A. - ULLMAN J. - ZEBER D. - ZHILYAEV M. Distributed Differential Privacy via Shuffling. In: ISHAI Y. - RIJMEN V. (eds) Advances in Cryptology – EUROCRYPT 2019. Lecture Notes in Computer Science, vol. 11476. Springer, Cham. Dostupné na: https://doi.org/10.1007/978-3-030-17653-2_13.

3. Vybrané právne úpravy navrhované Európskou komisiou vo vzťahu k šíreniu dezinformácií a správaniu sa online platforiem

Šírenie dezinformácií a opatrenia, ktoré limitujú alebo moderujú prejavy na internete, nevyhnutne narážajú na mantinely základných ľudských práv a slobôd. Systém ochrany základných ľudských práv a slobôd je v SR trojstupňový. Prvým stupňom je právna úprava v II. hlave Ústavy SR, ktorá obsahuje základné práva a slobody. Ako súčasť EÚ je Slovenská republika viazaná ľudsko-právnou úpravou v primárnom práve EÚ, ktorého súčasťou je aj Charta základných práv Európskej únie, čo tvorí druhý stupeň. Zároveň, v rámci tretieho stupňa ochrany je SR signatárom viacerých medzinárodnoprávných záväzkov, okrem iného aj Dohovoru Rady Európy o ochrane ľudských práv a základných slobôd.⁸¹

Pri regulácii dezinformácií na online platformách je potrebné zvýrazniť, že môže dôjsť ku zásahom do základných práv a slobôd tak užívateľov (jednotlivcov), ako aj samotných platforiem. Z hľadiska jednotlivcov môžeme diskutovať o zásahoch do slobody prejavu, práva na informácie, práva na súkromie a ochrany osobných údajov či práva na spravodlivý proces. Činnosť online platforiem na druhej strane podporuje právo šíriť informácie, právo vlastniť majetok či právo na podnikanie (Obrázok č. 7).



Obrázok č. 7

Šírenie dezinformácií na online platformách a potenciálne zásahy do základných ľudských práv a slobôd. Tmavšiu farbu majú práva a slobody, ktoré sa týkajú jednotlivcov. Svetlejší odtieň sa týka online platforiem.

Zdroj: Vlastná tvorba autorov.

⁸¹ Pozri napr. HUSOVEC, M. - MESARČÍK, M. Ľudské práva ako všeobecný limit technológií. In HUSOVEC, M. - MESARČÍK, M. - ANDRAŠKO, J. Právo informačných a komunikačných technológií I. TINCT, 2021, s. 66 a nasl.

Všetky vyššie uvedené právne akty obsahujú výpočet základných práv a slobôd, pričom zásah do nich musí byť v každom prípade vyvážený, nevyhnutný a proporcionálny. Jedným z najčastejšie spomínaných základných práv v súvislosti s dezinformáciami je **sloboda prejavu**. Sloboda prejavu je pilierom demokratických spoločností a nevyhnutnosťou pre spoločenský dialóg.⁸² Verejná diskusia a demokratické procesy sú však v súčasnosti vystavené častokrát zámernému šíreniu dezinformácií,⁸³ preto je na mieste uvažovať, ako vhodne túto problematiku regulovať a obmedziť slobodu prejavu.

Sloboda prejavu nechráni iba jednotlivcov, ale aj poskytovateľov služieb. Typické mocenské zásahy do tohto práva v kontexte nových technológií predstavujú odstraňovanie obsahu na internete alebo blokovanie webstránok, predovšetkým prostredníctvom súdnej moci.

Sloboda prejavu ale nie je jediným základným právom, ktoré šírenie dezinformácií ovplyvňuje. Diskutovať je potrebné aj zásahy do **práva na súkromie a práva na ochranu osobných údajov**⁸⁴ z toho dôvodu, že dezinformácie môžu významne ovplyvniť súkromie jednotlivca a zároveň ich odhaľovanie môže zahŕňať automatizovanú analýzu rôznych typov osobných údajov. Právne požiadavky pre spracúvanie osobných údajov sú v práve EÚ unifikované v podobe nariadenia GDPR.

Európsky súd pre ľudské práva veľmi často rieši prípady súvisiace so šírením informácií a slobodou prejavu. Známe sú predovšetkým rozhodnutia voči Poľsku, ktoré sa týkajú šírenia dezinformácií v rámci volebných kampaní, kde ESLP kritizoval poľské volebné zákony. V prípade *Brzeziński proti Poľsku* súd judikoval porušenie slobody prejavu z dôvodu, že súdy klasifikovali politické výroky ako lži bez ďalšieho skúmania. Podobné konštatovanie pri skúmaní pravdivosti informácie prezentoval súd aj v prípade *Kwiecień proti Poľsku*. V neposlednom rade, súd v prípade *Kita proti Poľsku* opätovne judikoval porušenie slobody prejavu, keďže poľské súdy nepostupovali podľa objektívnych kritérií a všetky výroky sťažovateľa klasifikovali ako nepravdivé bez skúmania faktických okolností.⁸⁵

⁸² Európska komisia. SPOLOČNÉ OZNÁMENIE EURÓPSKEMU PARLAMENTU, EURÓPSKEJ RADE, RADE, EURÓPSKEMU HOSPODÁRSKEMU A SOCIÁLNEMU VÝBORU A VÝBORU REGIÓNOV Akčný plán proti dezinformáciám. JOIN/2018/36 final. Dostupné na: <https://eur-lex.europa.eu/legal-content/SK/TXT/?uri=CELEX%3A52018JC0036>.

⁸³ Tamže.

⁸⁴ Charta základných práv EÚ a aj Ústava SR výslovne rozlišuje medzi právom na súkromie a právom na ochranu osobných údajov a upravuje ich osobitne. Pozri bližšie FUSTER, G. - HIJMANS, H. The EU rights to privacy and personal data protection: 20 years in 10 questions. Dostupné na: https://brusselsprivacyhub.eu/events/20190513.Working_Paper_Gonza%CC%81lez_Fuster_Hijmans.pdf.

⁸⁵ Bližšie ku judikatúre Európskeho súdu pre ľudské práva a slobode prejavu pozri štúdiu Európskeho parlamentu. The fight against disinformation and the right to freedom of expression. Dostupné na: [https://www.europarl.europa.eu/RegData/etudes/STUD/2021/695445/IPOL_STU\(2021\)695445_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2021/695445/IPOL_STU(2021)695445_EN.pdf).

Podobne, **Ústavný súd Slovenskej republiky** pri analýze ustanovení trestného zákona konštatoval problematiku klasifikácie informácií ako nepravdivých: "Rozhodovanie o verbálnych nenávisťných trestných činoch je veľmi delikátne. Pri krádeži je vec jasná. V podstate nie je čo ústavne vyvažovať. Ale pri nenávisťných trestných činoch sú v hre tak vysoké hodnoty ako sloboda prejavu na jednej strane a spoločenská kohézia, resp. spoločnosť štátu vôbec. Nenávisťné trestné činy sú jediné právne normy, v ktorých môže byť trestaný názor často nezmyselný, ba hlúpy, ale názor.(...) Celú aktuálnu situáciu ešte komplikuje špecifické prostredie internetu. Akékoľvek rozhodnutie ústavného súdu nemôže byť spoločensky ideálne, pretože oslabuje jednu z cenných ústavných hodnôt."⁸⁶

Z pohľadu užívateľov taktiež za zmienku stojí právo na spravodlivý súdny proces. Spravodlivý proces v zmysle ľudskoprávnej úpravy neznamena iba právo na prístup k súdu, prípadne riešenie sporu v čo najkratšom čase. Čoraz viac judikatúra, ale aj prax predmetné právo chápe aj optikou požiadavky spravodlivého procesu (*due process*), v čo najširšom slova zmysle. Z tohto dôvodu je nevyhnutné užívateľom zabezpečiť transparentnosť a nástroje na ochranu ich práv pri automatizovanom rozhodovaní, vrátane využívania umelej inteligencie na online platformách. Literatúra v zmysle práva na spravodlivý proces a algoritmického rozhodovania akcentuje predovšetkým transparentnosť, zodpovednosť a napadnuteľnosť takýchto rozhodnutí.⁸⁷

Ako zároveň uznávajú aj strategické dokumenty na úrovni EÚ pre boj s dezinformáciami, ich šírenie prostredníctvom online platforiem vrátane sociálnych médií je kľúčovým prvkom. Regulácia pravidiel pre podnikanie a činnosť súkromných entít nevyhnutne kolide s právom na podnikanie či právom vlastníť majetok, vrátane ochrany práv duševného vlastníctva. Z tohto dôvodu je nutné spomenúť právne úpravy, ktoré sa týkajú **zodpovednosti online platforiem**.

Všeobecná právna úprava zodpovednosti za cudzí obsah pri službách informačnej spoločnosti sa nachádza v **Smernici o elektronickom obchode**, ktorá podľa typu prevádzkovateľa upravuje podmienky, za ktorých nie je za obsah na svojej platforme zodpovedný (tzv. bezpečný prístav). V zmysle súčasnej právnej úpravy platí princíp *notice-to-takedown*, čo prakticky znamená, že ak platforma nemá vedomosť o problematickom obsahu, nie je zaň zodpovedná.⁸⁸ Ak sa platforma nemôže na bezpečný prístav spoľahnúť (napríklad ak jej niekto nelegálny obsah nahlási), môže byť braná na zodpovednosť prostredníctvom národného práva. V Slovenskej republike predovšetkým prostredníctvom občianskoprávnej zodpovednosti.⁸⁹

⁸⁶ Nález Ústavného súdu SR. PL. ÚS 5/2017-117. Bod, 71.2.

⁸⁷ Pozri viac napr. FROSIO, G. - GEIGER, CH. Taking Fundamental Rights Seriously in the Digital Services Act's Platform Liability Regime. European Law Journal (October 2021, Forthcoming), s. 27 - 28. Dostupné na: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3747756.

⁸⁸ Podrobnejšie pozri HUSOVEC, M. Zodpovednosť na internete podľa českého a slovenského práva. CZ-NIC, 2014. Dostupné na: https://knihy.nic.cz/files/edice/zodpovednost_na_internete.pdf.

⁸⁹ Teoretické odôvodnenie využívania daného modelu zodpovednosti je bližšie vysvetlené v literatúre, napr. FROSIO, G., - HUSOVEC, M. Accountability and Responsibility of Online Intermediaries in FROSIO, G. (ed). The Oxford Handbook of Online Intermediary Liability, s. 613 - 630.

Zároveň podnikanie súvisí aj s možnosťou svoje výrobky alebo služby propagovať prostredníctvom reklamy. Ako je uvedené v časti 2 tejto štúdie, do popredia sa dostali automatizované nástroje určené na profilovanie a využívanie umelej inteligencie pre personalizovanú reklamu. Práve tieto otázky reguluje oblasť ochrany osobných údajov, reklamy či ochrany spotrebiteľa.

Vzhľadom na ekonomickú silu a postavenie online platforiem na trhu ďalšie nástroje na ich reguláciu a obmedzovanie činnosti predstavuje **súťažné právo**, ktoré špecifickým spôsobom obmedzuje právo podnikať. Túto oblasť vo vzťahu ku veľkým online platformám bude regulovať tzv. Akt o digitálnych trhoch.⁹⁰

V súvislosti so šírením dezinformácií v súčasnosti EÚ predstavila balík legislatívnych aktov, ktoré majú výrazne pomôcť pri boji s dezinformáciami. Jednou zo zásadných zmien sú nové pravidlá pre reguláciu online platforiem v podobe Aktu o digitálnych službách (Digital Services Act, **DSA**), ktoré zároveň revidujú režim zodpovednosti v Smernici o elektronickom obchode.

Nakoľko legislatívny proces z objektívnych dôvodov trvá roky, EÚ v kontexte boja s dezinformáciami predstavila **Kódex nakladania s dezinformáciami**, ktorý by mal tvoriť akýsi "predvoj" pred prijatím DSA. Zároveň ide o nástroj, ktorý sa zaviazali dodržiavať najväčšie online platformy. Kódex upravuje špecifické pravidlá pre kontrolu umiestňovania reklamy; politickú reklamu a reklamu zameranú na aktuálne problémy; integritu služieb; posilnenie postavenia spotrebiteľov a výskumnej obce. Dodržiavanie pravidiel podlieha dohľadu samotných online platforiem a taktiež Európskej komisie. Samotná komisia predstavila v lete 2021 usmernenie pre posilnenie kódexu.⁹¹ V súlade s týmto usmernením by mala byť vypracovaná aj aktuálnejšia verzia kódexu. Monitorovanie súladu s kódexom ukazuje, že nie všetky online platformy efektívne pretavujú požiadavky do praxe.⁹²

Na jar 2021 bol predstavený návrh regulácie umelej inteligencie v podobe aktu o umelej inteligencii (Artificial Intelligence Act, **AIA**). V kontexte záverov časti 2 štúdie je preto nevyhnutné, že pravidlá pre vývoj a používanie umelej inteligencie sú v boji proti šíreniu dezinformácií nemenej dôležité.

Práve z dôvodu aktuálnosti predstavenia nových právnych úprav a vzhľadom na to, že šírenie dezinformácií je dnes zvýraznené využitím umelej inteligencie, sa budeme bližšie venovať inštitútom, ktoré obsahuje DSA a AIA.

⁹⁰ Návrh nariadenia Európskeho parlamentu a Rady o súťažeschopných a spravodlivých trhoch digitálneho sektora (akt o digitálnych trhoch). COM/2020/842 final. Dostupné na: <https://eur-lex.europa.eu/legal-content/SK/TXT/?uri=COM%3A2020%3A842%3AFIN#>.

⁹¹ Európska komisia. OZNÁMENIE KOMISIE EURÓPSKEMU PARLAMENTU, RADE, EURÓPSKEMU HOSPODÁRSKEMU A SOCIÁLNEMU VÝBORU A VÝBORU REGIÓNOV Usmernenie Európskej komisie k posilneniu Kódexu postupov proti šíreniu dezinformácií. COM/2021/262 final. Dostupné na: <https://eur-lex.europa.eu/legal-content/SK/TXT/?uri=CELEX%3A52021DC0262#>.

⁹² Napríklad ERGA. ERGA Report on disinformation: Assessment of the implementation of the Code of Practice. Dostupné na: <http://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf>.

3.1 Návrh nariadenia o umelej inteligencii (AIA)

V apríli 2021 Európska komisia prezentovala návrh prvej komplexnej regulácie umelej inteligencie v EÚ - návrh nariadenia o umelej inteligencii, známy pod skratkou AIA (Artificial Intelligence Act). Komisia zdôvodňuje prijatie legislatívy viacerými dôvodmi, ktoré menuje a analyzuje v tzv. posúdení vplyvu.⁹³ Medzi dôvody prijatia regulácie umelej inteligencie Komisia zaradila zvýšené riziko pre bezpečnosť jednotlivcov, riziká pre základné práva a slobody či nedostatok právomoci orgánov verejnej moci pri dohľade nad dodržiavaním pravidiel. Komisia zvýraznila aj právnu neistotu pri komplikovaných a fragmentovaných právnych úpravách, nedôveru v spoločnosti voči aplikáciám AI, ktorá môže v konečnom dôsledku brzdiť inovácie a ovplyvniť konkurencieschopnosť EÚ na globálnom trhu a hrozbu fragmentovaných národných právnych úprav, ktoré môžu predstavovať ohrozenie jednotného trhu EÚ.⁹⁴ Vyššie uvedené dôvody sú následne pretavené do konkrétnych právnych požiadaviek v návrhu AIA.

3.1.1 Všeobecná charakteristika

AIA predstavuje *produktovú horizontálnu legislatívu* na prístupe založenom na riziku. Horizontálnosť odzrkadľuje fakt, že AIA sa bude vzťahovať na systémy AI vo všeobecnosti a nie len v úzko vymedzených sektoroch.

Systém AI je v článku 3 ods. 1 AIA definovaný ako:

“softvér vyvinutý s jednou alebo viacerými technikami a prístupmi uvedenými v prílohe I, ktorý môže pre daný súbor cieľov vymedzených človekom vytvárať výstupy, ako je obsah, predpovede, odporúčania alebo rozhodnutia ovplyvňujúce prostredie, s ktorým sú v interakcii.”

Príloha č. 1 vymenúva zoznam techník a prístupov,⁹⁵ ktoré môže Európska komisia dynamicky dopĺňať s ohľadom na najnovší vývoj.

Definícia umelej inteligencie v rámci navrhovanej regulácie je problematická, pretože je koncipovaná veľmi široko, a bude zahŕňať obrovské množstvo aplikácií na báze AI.⁹⁶ V rámci našich verejne dostupných pripomienok sme akcentovali, že predmetná definícia by mala byť upravená takým spôsobom, aby techniky a prístupy uvedené v Prílohe č. 1 AIA boli integrálnou súčasťou systému AI a nie len „použitú“ pri jeho vývoji. Zároveň ich vplyv na generovanie výstupov by mal byť zohľadnený. Taktiež máme za to, že techniky a postupy v Prílohe č. 1 AIA

⁹³ Európska komisia. IMPACT ASSESSMENT Accompanying the Proposal for a Regulation of the European Parliament and of the Council LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS. SWD(2021) 84 final.

⁹⁴ Tamže, s. 13 - 28.

⁹⁵ Konkrétne menuje: “a) prístupy strojového učenia vrátane učenia s učiteľom, bez učiteľa a učenia posilňovaním pomocou širokého spektra metód vrátane hĺbkového učenia; b) prístupy založené na logike a poznatkoch vrátane reprezentácie poznatkov, induktívneho (logického) programovania, vedomostných základní, inferenčných a deduktívnych mechanizmov, (symbolického) uvažovania a expertných systémov; c) štatistické prístupy, bayesovský odhad, metódy vyhľadávania a optimalizácie.”

⁹⁶ Pozri napríklad Kempelen Institute of Intelligent Technologies. Stance on the Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. Dostupné na: https://kinit.sk/wp-content/uploads/2021/09/KINIT_Stance-of-AIA_Paper_2021_09.pdf. MULLER, C. - DIGNUM, V. ARTIFICIAL INTELLIGENCE ACT: ALLAI ANALYSIS & RECOMMENDATIONS. Dostupné na: <https://allai.nl/draft-ai-act-allai-analysis-and-recommendations/>.

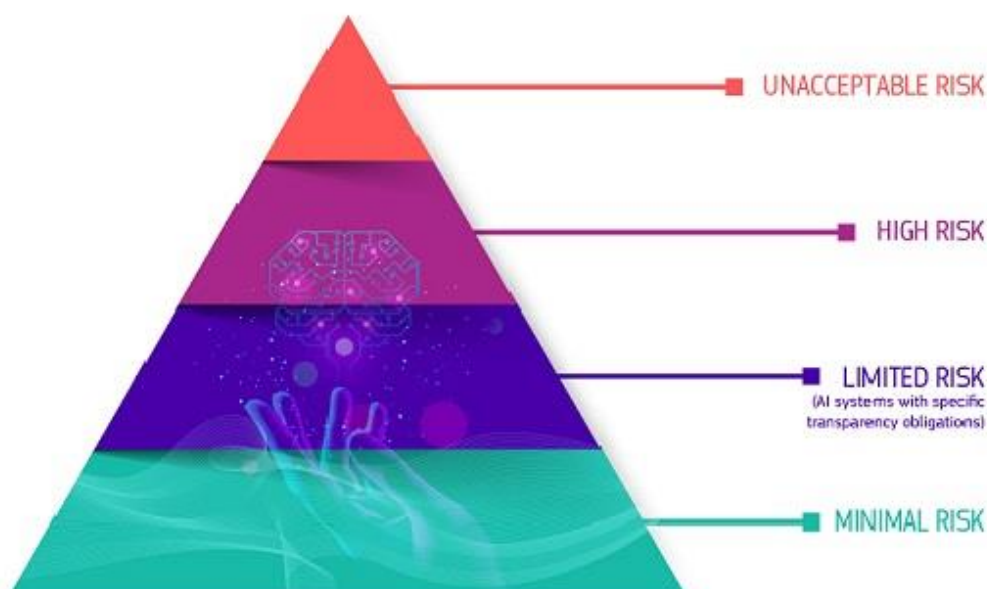
by nemali byť interpretované izolovane, ale iba pod podmienkou „inteligentného“ správania, ktorého kritériá (napr. schopnosť učenia sa) by mali byť definované expertnou komunitou.⁹⁷

AIA je koncipovaná na prístupe založenom na riziku. Prakticky to znamená, že kategorizuje systémy AI podľa miery rizika pre základné ľudské práva a slobody či bezpečnosť jednotlivcov. AIA rozlišuje (pozri obrázok č. 8):

1. neprijateľné riziko (zakázané praktiky)
2. vysoké riziko
3. nízke riziko
4. minimálne riziko.

Nemenej dôležitým faktorom je, že AIA je produktová regulácia. To znamená, že sa na systémy umelej inteligencie pozerá ako na produkt a na podmienky, ktoré musí spĺňať, aby mohol byť vyvíjaný a predstavený na jednotnom trhu EÚ. Neobsahuje výpočet konkrétnych práv jednotlivcov, ako to robí napríklad GDPR pri právach dotknutých osôb.

AIA z veľkej časti upravuje požiadavky pre systémy AI vysokého rizika. Tie budú musieť podstúpiť tzv. posudzovanie zhody (*conformity assessment*), kde AIA kladie viaceré požiadavky napríklad z pohľadu správy údajov, detekcie potenciálnych skreslení, technickej dokumentácie a ďalších oblastí. Posudzovanie zhody si budú musieť prevádzkovatelia systémov AI vo väčšine prípadov zabezpečiť vo vlastnej réžií. Následne môže prevádzkovateľ systémov AI disponovať EÚ vyhlásením o zhode, ktoré ho oprávňuje po registrácii v databáze EÚ pre samostatné vysokorizikové systémy umelej inteligencie uviesť daný systém na trh, kde podlieha požiadavkám monitorovania systému. V prípade, ak prevádzkovateľ systému AI už podlieha posúdeniu zhody podľa osobitných predpisov, pôvodné posúdenie zhody rozšíri o požiadavky vyplývajúce z AIA.



⁹⁷ Kempelen Institute of Intelligent Technologies. Stance on the Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. Dostupné na: https://kinit.sk/wp-content/uploads/2021/09/KINIT_Stance-of-AIA_Paper_2021_09.pdf.

AIA predpokladá dohľad nad dodržiavaním pravidiel prostredníctvom národných orgánov, ktoré si budú musieť členské štáty kreovať. Zároveň, podobne ako GDPR, AIA obsahuje vysoké pokuty za nedodržanie, pričom najväčšia správna pokuta môže byť uložená do až do výšky 30 000 000 EUR, alebo, ak je porušiteľom spoločnosť, až do výšky 6 % jej celkového svetového ročného obratu za predchádzajúci účtovný rok podľa toho, ktorá suma je vyššia.⁹⁹

3.1.2 Boj s dezinformáciami

AIA vo svojom texte neobsahuje priamu referenciu na dezinformácie. Totožné konštatovanie uvádzame aj pri posúdení vplyvu. To ale automaticky neznamená, že AIA sa šírenia dezinformácií s využitím systémov AI netýka. Práve naopak. Nakoľko ide o horizontálnu všeobecnú reguláciu, v rámci svojich inštitútov sa môže dotýkať aj online platforiem a šírenia dezinformácií. Nižšie pozornosť upriamime na otázky zakázaných praktík, systémov AI vysokého rizika a špecifické inštitúty v rámci posúdenia zhody a auditov.

3.1.2.1 Zakázané praktiky

AIA v článku 5 zakazuje tri praktiky v oblasti umelej inteligencie a jednu výrazne obmedzuje.

Konkrétne zakazuje:

- a) uvádzanie na trh, uvádzanie do prevádzky alebo používanie systému umelej inteligencie, ktorý **s cieľom podstatne narušiť správanie osoby využíva podprahové techniky mimo vedomia osoby** tak, že tejto alebo inej osobe spôsobí alebo by mohol spôsobiť fyzickú alebo psychickú ujmu;
- b) uvádzanie na trh, uvádzanie do prevádzky alebo používanie systému umelej inteligencie, ktorý **s cieľom podstatne narušiť správanie osoby patriacej do konkrétnej skupiny osôb využíva ktorúkoľvek zo zraniteľností tejto skupiny osôb vyplývajúcich z ich veku, fyzického alebo duševného postihnutia** tak, že tejto alebo inej osobe spôsobí alebo by mohol spôsobiť fyzickú alebo psychickú ujmu;
- c) uvádzanie na trh, uvádzanie do prevádzky alebo používanie systémov umelej inteligencie **orgánmi verejnej moci alebo v ich mene na účely hodnotenia alebo klasifikácie dôveryhodnosti fyzických osôb** počas určitého obdobia na základe ich spoločenského správania alebo známych či predpokladaných osobných alebo osobnostných charakteristík, pričom takto získané sociálne skóre vedie k jednému alebo obidvom z týchto výsledkov:
 - I. škodlivé alebo nepriaznivé zaobchádzanie s určitými fyzickými osobami alebo celými skupinami fyzických osôb v sociálnych kontextoch, ktoré nesúvisia s kontextmi, v ktorých boli údaje pôvodne generované alebo zhromaždené;

⁹⁸ European Commission. Regulatory framework proposal on artificial intelligence. Dostupné na: <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>.

⁹⁹ Pozri AIA, článok 71.

- II. škodlivé alebo nepriaznivé zaobchádzanie s určitými fyzickými osobami alebo celými skupinami fyzických osôb, ktoré je neodôvodnené alebo neprimerané ich spoločenskému správaniu alebo jeho závažnosti.”

Zároveň článok 5 ods. 1 písm. d) AIA v spojitosti s odsekom 2 výrazne obmedzuje používanie **systémov diaľkovej biometrickej identifikácie v reálnom čase vo verejne prístupných priestoroch na účely presadzovania práva**. AIA ustanovuje presné výnimky a manévrovací priestor členských štátov EÚ pre reguláciu daných výnimiek.

V kontexte využívania systémov AI a šírenia dezinformácií bude kľúčový výklad zakázaných praktík týkajúcich sa limitovania manipulácie,¹⁰⁰ či už s využitím podprahových techník mimo vedomia osoby alebo zraniteľností špecifickej skupiny osôb. Z hľadiska výkladov termínu manipulácia prezentovaných v literatúre návrh AIA reflektuje väčšinu požiadaviek. Vyžaduje úmysel (“s cieľom”), zneužitia zraniteľností (vek, psychická alebo fyzická porucha) či pôsobenie mimo prahu vedomia osoby. Absentuje však zásadný znak manipulácie, t. j. požiadavka podpory vlastných cieľov manipulujúcej osoby.¹⁰¹ Namiesto toho, AIA vyžaduje reálnu alebo potenciálnu psychickú či fyzickú ujmu.

Navyše, Recitál 16 AIA limituje využívanie týchto zákazov ešte výraznejšie a neodzrkadľuje dynamiku vzťahov pri využívaní systémov AI:

“Takýto úmysel **nemožno predpokladať, ak ovplyvnenie ľudského správania vyplýva z faktorov nesúvisiacich** so systémom umelej inteligencie, ktoré sú **mimo kontroly poskytovateľa alebo používateľa.**”

Prakticky to môže znamenať, že odporúčacie systémy na online platformách, ktoré uprednostňujú dezinformácie, nebudú pod zakázanú praktiku spadať. Tento záver prezentujeme hlavne z toho dôvodu, že súčasný návrh AIA nevyžaduje splnenie požiadavky manipulácie s cieľom podporiť svoje ciele. Bolo by veľmi ťažko preukázateľné, že všetky dezinformácie *en bloc* môžu potenciálne spôsobiť jednotlivcom ujmu. Zároveň problémom je, že tieto zakázané praktiky absolútne nezohľadňujú kontext a v tejto chvíli nepoznáme presnú interpretáciu daných praktík do budúcnosti.

3.1.2.2 Vysokorizikové systémy AI

Ako sme už uviedli vyššie, väčšina ustanovení v návrhu AIA sa bude vzťahovať na systémy AI vysokého rizika. Systém AI môže byť klasifikovaný ako systém vysokého rizika dvoma spôsobmi. Prvým spôsobom je odkaz na špecifickú produktovú reguláciu, ktorá už obsahuje určité požiadavky na systémy AI ako bezpečnostné komponenty.¹⁰² Druhým spôsobom sú prípady samostatných systémov AI, ktoré sú uvedené v Prílohe III AIA.¹⁰³

¹⁰⁰ AIA, Recitál 15: “Využívanie umelej inteligencie má síce mnoho výhod, túto technológiu však možno zneužiť a môže sa stať zdrojom nových a výkonných nástrojov **umožňujúcich manipulatívne a zneužívajúce praktiky** a praktiky v oblasti sociálnej kontroly.”

¹⁰¹ VAELE, M. - BORGESIU, F.Z. Demystifying the Draft EU Artificial Intelligence Act. Forthcoming in 2021 22(4) Computer Law Review International.

¹⁰² AIA, článok 6 ods. 1.

¹⁰³ AIA, článok 6 ods. 2.

Európska komisia v Prílohe III uvádza zoznam oblastí a aplikácií systémov AI, ktoré považuje za systémy AI vysokého rizika:

- Biometrická identifikácia a kategorizácia fyzických osôb;
- Riadenie a prevádzka kritickej infraštruktúry;
- Vzdelávanie a odborná príprava;
- Zamestnanosť, riadenie pracovníkov a prístup k samostatnej zárobkovej činnosti;
- Prístup k základným súkromným a verejným službám a dávkam a ich využívanie;
- Presadzovanie práva;
- Migrácia, azyl a riadenie kontroly hraníc.

Bližší pohľad na oblasti a konkrétne aplikácie systémov AI vysokého rizika prezrádza, že klasické modely ekonomiky pozornosti (*attention economy*) online platforiem charakterizované v časti 2 tejto štúdie nie sú v rámci systémov AI vysokého rizika podľa Prílohy III zahrnuté. Nie je možné tieto systémy subsumovať ani pod jednu z oblastí alebo aplikácií.

Tento fakt považujeme za zásadný nedostatok AIA, ktorý sme komunikovali aj verejne na inom mieste.¹⁰⁴ Navrhujeme oblasť ekonomiky pozornosti zaradiť medzi vysokorizikové systémy umelej inteligencie podobne, ako je to v prípade využívania biometrie, kde je časť zakázaná (obmedzená) a časť podlieha požiadavkám na systémy AI vysokého rizika.

3.1.2.3 Požiadavky na systémy AI vysokého rizika a monitoring po uvedení na trh

V prípade zaradenia AI systémov, ktoré využívajú online platformy a ktoré sa môžu podieľať na šírení dezinformácií, medzi vysokorizikové systémy AI, by sa na ne vzťahovalo niekoľko zásadných požiadaviek. V súvislosti so systémami AI vysokého rizika totiž AIA ustanovuje viacero kľúčových požiadaviek, ktoré musí prevádzkovateľ systému AI zohľadniť pri posudzovaní zhody.

Jednou z nich je aj systém riadenia rizík podľa článku 9 AIA, ktorý zahŕňa “nepretržitý opakovaný proces počas celého životného cyklu vysokorizikového systému umelej inteligencie, ktorý si vyžaduje pravidelné systematické aktualizovanie.” AIA zároveň priamo uvádza minimálne požiadavky na systém riadenia rizík, ktoré sú ale pomerne všeobecné a musia byť prispôbené konkrétnemu prípadu.

Z hľadiska správy údajov obsahuje AIA dôležitú požiadavku pre súbory tréningových, validačných a testovacích údajov, a síce musia platiť primerané postupy správy a riadenia údajov, ktoré vyžadujú preskúmania z hľadiska možných skreslení (*bias detection*).¹⁰⁵

¹⁰⁴ Kempelen Institute of Intelligent Technologies. Stance on the Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. Dostupné na: https://kinit.sk/wp-content/uploads/2021/09/KINIT_Stance-of-AIA_Paper_2021_09.pdf.

¹⁰⁵ AIA, článok 10 ods. 2 písm. f).

Nemenej kľúčové požiadavky sa týkajú transparentnosti voči biznisovým užívateľom¹⁰⁶ a ľudskému dohľadu.¹⁰⁷

Po uvedení systému AI vysokého rizika na trh platí požiadavka zavedenia a zdokumentovania systému monitorovania spôsobom, ktorý je primeraný povahe technológii umelej inteligencie a rizikám vysokorizikového systému umelej inteligencie.¹⁰⁸ Predmetná požiadavka je ale formulovaná veľmi všeobecne a nemožno z nej vyčítať konkrétne parametre potenciálnych verejne dostupných správ alebo auditu.

3.2 Návrh nariadenia o digitálnych službách (DSA)

Na konci roka 2020 predstavila Európska komisia návrh nariadenia o digitálnych službách známeho pod skratkou DSA (**D**igital **S**ervices **A**ct). Základným motívom bola revízia dvadsať rokov platného právneho rámca zodpovednosti online platforiem v podobe Smernice o elektronickom obchode. Zároveň zákonodarca reflektoval vývoj digitálneho trhu a narastajúcej ekonomickej sily online platforiem, čo podnietilo diskusie o možných rizikách a regulácii.

Komisia identifikovala tri konkrétne dôvody, ktoré viedli k navrhovanej legislatíve:

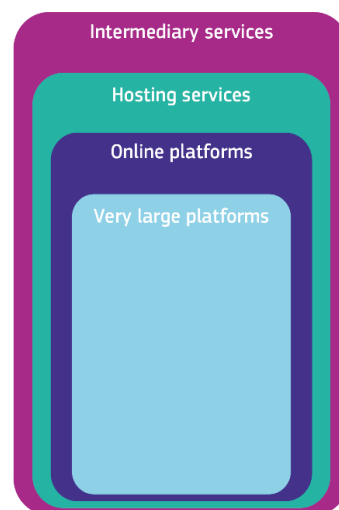
- V prvom rade ide o nárast rizík v súvislosti s používaním a správaním online platforiem v podobe sociálnych a ekonomických rizík a možnej ujmy pre jednotlivcov a ich základné práva a slobody.
- Druhým dôvodom je nedostatočný dohľad a spolupráca pri digitálnych službách, čo spôsobuje eróziu jednotného trhu EÚ.
- Tretím dôvodom je odstránenie bariér pre menšie spoločnosti pri poskytovaní digitálnych služieb vzhľadom na pozíciu veľkých online platforiem.¹⁰⁹ Tieto dôvody sú následne prenesené do konkrétnych požiadaviek v DSA.

3.2.1 Všeobecná charakteristika

Návrh DSA z hľadiska osobnej pôsobnosti rozlišuje štyroch aktérov, na ktorých sa vzťahujú konkrétne právne požiadavky. V zásade ide o niekoľko množín vzťahov.

DSA rozlišuje (pozri obrázok č. 9):

- Sprostredkovateľské služby;
- Hostingové služby;
- Online platformy; a



¹⁰⁶ AIA, článok 13.

¹⁰⁷ AIA, článok 14.

¹⁰⁸ AIA, článok 61.

¹⁰⁹ Európska komisia. IMPACT ASSESSMENT Accompanying the document PROPOSAL FOR A REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC. SWD(2020) 348 final.

- Veľmi veľké online platformy.¹¹⁰

Najväčšou množinou aktérov sú **sprostredkovateľské služby**, ktoré návrh DSA delí na služby typu obvyčajný prenos, kešing a hosting.¹¹¹ Služby typu obvyčajný prenos zahŕňajú poskytovateľov internetového pripojenia alebo prevádzkovateľov otvorenej wifi siete. Podstatou služieb kešingu je dočasné a prechodné uloženie informácií s cieľom zefektívnenia služby.

Dôležitou množinou z hľadiska šírenia dezinformácií sú služby typu **hosting**. V zásade ide o služby webhostingu, cloudových služieb, úložísk, sociálnych sietí alebo online inzercii.

Podmnožinou služieb hosting sú **online platformy**, ktoré DSA definuje ako

„poskytovateľ hostingovej služby, ktorý na žiadosť príjemcu služby uchováva a verejne šíri informácie, pokiaľ táto činnosť nie je nevýznamným a čisto vedľajším prvkom inej služby, ktorý z objektívnych a technických dôvodov nemožno použiť bez tejto inej služby, a jeho začlenenie do inej služby nie je prostriedkom na obchádzanie uplatniteľnosti tohto nariadenia.“¹¹²

Medzi online platformy môžeme typicky zaradiť sociálne média, online trhoviská či obchody s aplikáciami. Na prvý pohľad sa zdá, že daná definícia nebude zahŕňať služby na posielanie správ ako WhatsApp alebo Telegram. Tento záver ale bude závisieť od výkladu pojmu „verejného šírenia informácií“ a môže sa v procese doladovania legislatívy ešte zmeniť.

Špecifická pozornosť a právne požiadavky sú zamerané aj na poslednú množinu aktérov - **veľmi veľké online platformy**. Tieto priamo definované v DSA nie sú, avšak návrh nariadenia obsahuje návod na klasifikáciu. V zmysle článku 25 ods. 1 DSA veľkými online platformami možno rozumieť tých sprostredkovateľov, „ktorí svoje služby poskytujú priemernému mesačnému počtu aktívnych príjemcov služby v Únii, ktorý sa rovná 45 miliónom alebo je vyšší a ktorý je vypočítaný v súlade s metodikou“ stanovenou v DSA.

Požiadavky kladené DSA na vyššie uvedených aktérov sa líšia a je preto vždy dôležité správne klasifikovať sprostredkovateľa služieb. Nižšie uvádzame stručný prehľad aktérov a vybraných povinností v zmysle DSA (tabuľka č.3).¹¹³

Obrázok č. 9

Množina entít v rámci pôsobnosti DSA.

Zdroj: Webová stránka Európskej komisie.¹¹⁰

¹¹⁰ Európska komisia. The Digital Services Act: ensuring a safe and accountable online environment. Dostupné na: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en.

¹¹¹ DSA, článok 2 písm. f):

„sprostredkovateľská služba“ je jedna z týchto služieb:

- služba „**obvyčajný prenos**“, ktorá pozostáva z prenosu informácií poskytovaných príjemcom služby v komunikačnej sieti alebo z poskytovania prístupu ku komunikačnej sieti,
- služba „**kešing**“, ktorá pozostáva z prenosu informácií poskytovaných príjemcom služby v komunikačnej sieti, pri ktorom sa tieto informácie automaticky, dočasne a prechodne uchovávajú, a to výlučne na účely zefektívnenia ďalšieho prenosu informácií k iným príjemcom na ich žiadosť,
- služba „**hosting**“, ktorá pozostáva z uchovávaní informácií poskytovaných príjemcom služby na jeho žiadosť.“

¹¹² DSA, článok 2 písm. h).

¹¹³ Európska komisia. The Digital Services Act: ensuring a safe and accountable online environment. Dostupné na: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en.

Tabuľka č. 3
Povinnosti pre rôzne subjekty z pohľadu DSA.
Zdroj: DSA.

#	SPROSTREDKO VATEĽSKÁ SLUŽBA	HOSTING	ONLINE PLATFORMA	VEĽMI VEĽKÁ ONLINE PLATFORMA
Transparentnosť a reportovanie	✓	✓	✓	✓
Požiadavky na všeobecné obchodné podmienky	✓	✓	✓	✓
Spolupráca s národnými orgánmi	✓	✓	✓	✓
Určenie kontaktného miesta (zástupcu)	✓	✓	✓	✓
Informačné povinnosti		✓	✓	✓
Mechanizmy pre vybavovanie sťažností a mimosúdne riešenie sporov			✓	✓
Dôveryhodní nahlasovatelia			✓	✓
Opatrenia na predchádzanie zneužívajúcich notifikácií			✓	✓
Sledovateľnosť obchodníkov			✓	✓
Transparentnosť reklamy			✓	✓

#	SPROSTREDKO VATEĽSKÁ SLUŽBA	HOSTING	ONLINE PLATFORMA	VEĽMI VEĽKÁ ONLINE PLATFORMA
Manažment rizík a určenie zodpovednej osoby				✓
Externé audity a verejná zodpovednosť (accountability)				✓
Transparentnosť odporúčacích systémov				✓
Spolupráca pri riešení krízových situácií				✓

Podobne ako aj v prípade AIA, DSA predpokladá určenie kompetentného orgánu, ktorý bude dozerať na plnenie požiadaviek ustanovených v DSA v jednotlivých členských štátoch. Konkrétne sankcie si budú môcť určiť členské štáty vo vlastnej réžii. DSA ale predpokladá, že maximálna výška sankcií uložených za nedodržanie povinností stanovených v DSA nepresiahne 6% ročného príjmu alebo obratu dotknutého poskytovateľa sprostredkovateľských služieb.¹¹⁴ Voči veľmi veľkým online platformám bude môcť konať aj samotná Európska komisia.¹¹⁵

3.2.2 Boj s dezinformáciami

Európska komisia sa v súlade so závermi prezentovanými v dokumente “Boj proti dezinformáciám na internete: európsky prístup” rozhodla neprijímať všeobecnú reguláciu smerovanú výlučne na boj s dezinformáciami.¹¹⁶ Namiesto toho predstavila “mäkší” mechanizmus v podobe vyššie spomínaného Kódexu nakladania s dezinformáciami. Zároveň samotné DSA čiastočne reflektuje požiadavku väčšej zodpovednosti online platforiem za šírený obsah s cieľom brániť úmyselnej manipulácii služieb ich užívateľmi prostredníctvom falošných účtov alebo botov.¹¹⁷

Boj s dezinformáciami je v kontexte DSA výslovne spomínaný v posúdení vplyvu a zároveň aj v recitálovej časti nariadenia. Posúdenie vplyvu uvádza šírenie dezinformácií ako jednu z aktivít, ktorá síce nemusí striktné spadať pod definíciu nezákonného obsahu, ale zároveň predstavuje obrovské riziko z hľadiska robenia informovaných rozhodnutí a otvorenosti politických

¹¹⁴ DSA, článok 42.

¹¹⁵ DSA, článok 51 a nasl.

¹¹⁶ Európska komisia. OZNÁMENIE KOMISIE EURÓPSKEMU PARLAMENTU, RADE, EURÓPSKEMU HOSPODÁRSKEMU A SOCIÁLNEMU VÝBORU A VÝBORU REGIÓNOV Boj proti dezinformáciám na internete: európsky prístup. COM/2018/236 final. Dostupné na: <https://eur-lex.europa.eu/legal-content/SK/TXT/?uri=CELEX%3A52018DC0236>.

¹¹⁷ HOBOKEN VAN, J. - Ó FATHAIGH, R. Regulating Disinformation in Europe: Implications for Speech and Privacy. UC Irvine Journal of International, Transnational, and Comparative Law. Volume 6 Symposium: The Transnational Legal Ordering of Privacy and Speech. Article 3, 2021, s. 15 - 16.

procesov. Práve šírenie dezinformácií môže byť zvýraznené a násobené prostredníctvom algoritmov online platforiem¹¹⁸ či organizovaných aktivít botov, ktoré umelo navýšia zaujímavosť či dosah obsahu a prispievajú k šíreniu dezinformácií.¹¹⁹ Predmetné konštatovanie osobitne platí vo vzťahu k odporúčacím systémom.¹²⁰

Vyššie uvedené riziká zvýrazňuje aj nezáväzný text recitálov DSA vo vzťahu k reklamným systémom veľkých online platforiem a sprístupňovaniu ich archívov.¹²¹ DSA zdôrazňuje rolu **kódexov správania** pri šírení dezinformácií, pričom tieto kódexy by sa mali veľké online platformy zaviazat' dodržiavať:

“Ďalšou oblasťou, ktorá si zaslúži pozornosť, sú možné negatívne vplyvy systémových rizík na spoločnosť a demokraciu, ako sú **dezinformácie** alebo manipulatívne a nekalé činnosti. Patria sem koordinované operácie zamerané na šírenie informácií vrátane **dezinformácií, ako je používanie botov alebo falošných účtov na vytváranie falošných alebo zavádzajúcich informácií, niekedy s cieľom dosiahnuť ekonomický zisk, ktoré sú obzvlášť škodlivé pre zraniteľných príjemcov služieb**, napríklad pre deti. Pokiaľ ide o tieto oblasti, možno za vhodné opatrenie na zmiernenie rizika považovať dodržiavanie príslušného **kódexu správania** zo strany veľmi veľkej online platformy. Ak online platforma bez riadneho vysvetlenia odmietne výzvu Komisie, aby sa podieľala na uplatňovaní takéhoto kódexu správania, mohlo by sa to prípadne zohľadniť pri rozhodovaní o tom, či online platforma porušila povinnosti stanovené v tomto nariadení.”¹²²

Otázka dezinformácií sa spomína aj v kontexte vypracovania krízových protokolov v prípade mimoriadnych situácií (napríklad pandémie alebo prírodná katastrofa), kde sa predpokladá spolupráca veľkých online platforiem s kompetentnými autoritami.¹²³

Na záver tejto state je potrebné zvýrazniť, že DSA výrazne reguluje kroky online platforiem súvisiace s **nelegálnym obsahom**. DSA definuje nezákonný obsah nasledovne:

“akákoľvek informácia, ktorá sama osebe alebo tým, že odkazuje na nejakú činnosť vrátane predaja produktov alebo poskytovania služieb, **nie je v súlade s právnymi predpismi** Únie alebo niektorého členského štátu, a to bez ohľadu na presný predmet alebo povahu týchto právnych predpisov.”

Šírenie dezinformácií nemusí obsahovo nevyhnutne spĺňať definíciu nezákonného obsahu ako je napríklad detská pornografia alebo reklamy na psychotropné látky. Toto konštatovanie potvrdzuje aj Recitál 22 DSA: “Odstránenie obsahu alebo znemožnenie prístupu k nemu by sa malo uskutočniť v súlade so zásadou slobody prejavu.”

¹¹⁸ Európska komisia. IMPACT ASSESSMENT Accompanying the document PROPOSAL FOR A REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC. SWD(2020) 348 final, s. 16.

¹¹⁹ Tamže, s. 17.

¹²⁰ Tamže, s. 28.

¹²¹ DSA, Recitál 63.

¹²² DSA, Recitál 68.

¹²³ DSA, Recitál 71.

3.2.2.1 Zodpovednosť online platforiem

Základnou otázkou pri boji s dezinformáciami je zodpovednosť kľúčových aktérov v procese – platforiem, cez ktoré sa dezinformácie šíria najrýchlejšie. V tomto smere je potrebné diskutovať možnosti zodpovednosti za obsah, ktorý na platformy pridáva tretia strana (užívatelia). Optikou slovenského práva ide o situáciu, ktorá spočíva v spolupôsobení pri páchaní deliktov v podobe účasti, použitia osoby na spáchanie deliktu alebo nezakročenia na ochranu práv inej strany.¹²⁴ Kedy je osoba zodpovedná za cudzí obsah zverejnený na online platforme, upravuje dnes Smernica o elektronickom obchode a slovenský zákon o elektronickom obchode. Tento režim je zároveň aj predmetom úpravy v DSA.

Vzhľadom na zameranie tejto štúdie sa zameriame na otázky zodpovednosti za cudzí obsah služieb “hosting,” nakoľko práve do tohto typu služieb spadajú online platformy a veľmi veľké online platformy vrátane sociálnych sietí. Podobne ako pri Smernici o elektronickom obchode, DSA zachováva koncept tzv. “bezpečného prístavu”, a teda podmienok, ktoré ak hostingová služba splní, nie je zodpovedná za obsah tretích strán. DSA tieto podmienky upravuje v článku 5 ods. 1:

“poskytovateľ služby **nie je zodpovedný** za informácie uchovávané na žiadosť príjemcu služby pod podmienkou, že poskytovateľ:

- a) **nemá skutočnú vedomosť** o nezákonnej činnosti alebo nezákonnom obsahu a vzhľadom na nároky na náhradu škody si nie je vedomý skutočností alebo okolností, z ktorých by bolo zrejmé, že ide o nezákonnú činnosť alebo nezákonný obsah, alebo
- b) **po získaní takejto vedomosti alebo povedomia urýchlene koná**, aby nezákonný obsah odstránil alebo k nemu znemožnil prístup.”

Podmienkou na splnenie podmienok bezpečného prístavu a vyhnutiu sa zodpovednosti je pasivita prevádzkovateľa hostingovej služby. Filtrovanie a zobrazovanie konkrétneho obsahu konkrétnym užívateľom je považované za technickú a pasívnu záležitosť fungovania platforiem.¹²⁵ Toto konštatovanie je ale diskutovateľné v dobe aktivít platforiem pre udržanie pozornosti a algoritmického rozhodovanie o zobrazovanom obsahu.¹²⁶

Rakúsky najvyšší súd sa v prípade predloženom na riešenie aj Súdnemu dvoru Európskej Únie zaoberal otázkou, či online platformy ostávajú neutrálne, ak optimalizujú zobrazovanie svojho obsahu. Rakúsky najvyšší súd sa priklonil k interpretácii, že takéto konanie je tradičný obchodný model a štruktúrovanie vyhľadávania by stále malo byť považované za pasívne správanie platformy.¹²⁷

¹²⁴ HUSOVEC, M. Digitálny trh EÚ a zodpovednosť poskytovateľov služieb.. In HUSOVEC, M. - MESARČÍK, M. - ANDRAŠKO, J. Právo informačných a komunikačných technológií I. TINCT, 2021, s. 128 a nasl.

¹²⁵ BUITEN, M. The Digital Services Act: From Intermediary Liability to Platform Regulation. Working Paper. Dostupné na https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3876328, s. 15. Zhodne Európska komisia. IMPACT ASSESSMENT Accompanying the document PROPOSAL FOR A REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC. SWD(2020) 348 final, s. 159.

¹²⁶ BUITEN, M. The Digital Services Act: From Intermediary Liability to Platform Regulation. Working Paper. Dostupné na https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3876328, s. 15.

¹²⁷ Handelsgericht Wien [Higher Regional Court of Vienna] Puls 4 TV GmbH & Co. KG v YouTube LLC and Google Austria GmbH, prípad č. 4 R 119/18a, s. 10.

Zároveň prevádzkovateľ nesmie priamo a ani nepriamo navádzať na spáchanie deliktu, inak bezpečný prístav stráca a stáva sa zodpovedným.¹²⁸ V prípade, ak prevádzkovateľ získava skutočnú (napríklad na základe oznámenia) alebo konštruktívnu vedomosť (napríklad na základe svojej činnosti)¹²⁹ o nezákonnej činnosti alebo obsahu, je povinný konať s cieľom odstránenia alebo znemožnenia prístupu. DSA upravuje pomerne presný opis mechanizmu na oznamovanie nezákonného obsahu zo strany užívateľa či iného subjektu (tzv. *notice and action mechanism*). Ak tieto oznámenia spĺňajú nariadením ustanovené požiadavky, považujú sa za dostatočne presné na splnenie požiadavky “skutočnej vedomosti” o nelegálnom obsahu a prevádzkovatelia musia urýchlene konať a proti obsahu zakročiť.¹³⁰

V prípade online platforiem a veľmi veľkých online platforiem DSA upravuje aj nový inštitút dôveryhodných nahlasovateľov (*trusted flaggers*),¹³¹ ktorých oznámenia musia byť vybavené prednostne a bezodkladne. Status dôveryhodného nahlasovateľa môže entita (nie jednotliviec) získať od koordinátora digitálnych služieb po informovaní Európskej komisie.

Článok 6 DSA zároveň výslovne podporuje dobrovoľné vyšetrovania zo strany platforiem. Pri aktivitách zameraných na odhaľovanie, identifikáciu a odstraňovanie nezákonného obsahu alebo znemožnenie prístupu k nemu alebo prijímania potrebných opatrení na dosiahnutie súladu s právnymi požiadavkami sa môžu spoľahnúť na výnimku zo zodpovednosti uvedenej vyššie.¹³² Inými slovami, ak platforma vykonáva dobrovoľné šetrenie svojho obsahu, nemusí to automaticky z pohľadu práva znamenať nadobudnutie objektívnej vedomosti o nelegálnom obsahu. Platforma sa stále môže spoľahnúť na podmienky bezpečného prístavu.

Článok 7 DSA taktiež obsahuje zákaz všeobecnej povinnosti monitorovať obsah na svojej službe:

“Poskytovateľom sprostredkovateľských služieb sa **neukladá všeobecná povinnosť monitorovať informácie**, ktoré títo poskytovatelia prenášajú alebo uchovávajú, ani **povinnosť aktívne zisťovať** skutočnosti alebo okolnosti naznačujúce nezákonnú činnosť.”

Súdny dvor EÚ opakovane vo svojej judikatúre uvádza, že všeobecná povinnosť monitorovať obsah by pre prevádzkovateľov znamenala výrazný zásah do práva na podnikanie a v dlhodobom horizonte limity pre vývoj inovácií.¹³³ Azda najlepšie to vyjadril generálny advokát v prípade Youtube/Cyando v kontexte ochrany autorského práva.¹³⁴ Jeho závery je ale analogicky možné aplikovať aj na povinnosť kontroly akéhokoľvek obsahu pred publikovaním na platformách:

¹²⁸ DSA, článok 5 ods. 2: “Odsek 1 sa neuplatňuje, ak príjemca služby podlieha poskytovateľovi alebo je pod jeho kontrolou.”

¹²⁹ K tomu pozri napríklad rozhodnutie Súdneho dvora EÚ č. C-324/09 vo veci z 12. júla 2011 L'Oréal SA a iní proti eBay International AG a iní alebo spojené veci C-236/08 až C-238/08 z 23. marca 2010 Google.

¹³⁰ DSA, článok 14.

¹³¹ DSA, článok 19.

¹³² Ide o klauzulu tzv. Dobrého samaritána. Pozri viac KUCZERAWY, A. The Good Samaritan that wasn't: voluntary monitoring under the (draft) Digital Services Act. Dostupné na: <https://verfassungsblog.de/good-samaritan-dsa/>.

¹³³ Pozri napríklad rozhodnutie Súdneho dvora EÚ C-360/10 zo dňa 16. februára 2012 vo veci SABAM v Netlog.

¹³⁴ Rozhodnutie Súdneho dvora EÚ C-682/18 zo dňa 22. júna 2021 Frank Peterson proti Google LLC a i. a Elsevier Inc. proti Cyando AG.

“Požiadavka, aby prevádzkovatelia platforiem s cieľom vyhľadávania akéhokoľvek porušenia autorských práv **všeobecným a abstraktným spôsobom kontrolovali všetky súbory**, ktoré majú ich používatelia v úmysle zverejniť, ešte pred ich nahraním na internet, by pritom **predstavovala značné riziko zásahu do týchto rôznych základných práv**. Vzhľadom na potenciálne obrovské množstvo ukladaného obsahu by totiž jednak bolo nemožné manuálne vykonávať takúto predbežnú kontrolu, a jednak riziko spojené so zodpovednosťou by bolo pre týchto prevádzkovateľov neúmerné. V praxi by hrozilo, že najmenší prevádzkovatelia by túto zodpovednosť nezvládli, a tí, ktorí majú dostatočné zdroje, by boli nútení pristúpiť k všeobecnému filtrovaniu obsahu svojich používateľov bez súdneho preskúmania, čo by viedlo k následnému riziku „nadmerného odstraňovania“ tohto obsahu.”

3.2.2.2 Moderovanie obsahu

Nadväzujúc na systém nahlasovania nezákonného obsahu, DSA obsahuje dodatočné povinnosti týkajúce sa odôvodnenia odstránenia obsahu a interné a externé mechanizmy na riešenie sporov. V prípade, ak sa online platforma po preskúmaní rozhodne nezákonný obsah vymazať alebo znepriístupniť, DSA určuje presné požiadavky na odôvodnenie daného rozhodnutia príjemcom danej služby. Okrem iného musí dané odôvodnenie obsahovať vysvetlenie vymazania alebo nesprístupnenia nezákonného obsahu a či bol tento krok vykonaný automatizovaným spôsobom.¹³⁵ DSA tak nepriamo reflektuje požiadavku ľudského dohľadu pri mechanizme preskúmania rozhodnutí o vymazaní alebo nesprístupnení obsahu. Všetky rozhodnutia online platforiem budú zverejnené v registri spravovanom Európskou komisiou.

Voči rozhodnutiam online platforiem poskytuje DSA viaceré mechanizmy právnej obrany. V prvom rade ide o interný systém vybavovania sťažností. Príjemca služby má možnosť podať sťažnosť v prípade, ak platforma vyhodnotila jeho aktivitu ako šírenie nezákonného obsahu alebo v nesúlade so všeobecnými obchodnými podmienkami, čo má za následok rozhodnutie o vymazaní obsahu, (čiastočnom) pozastavení poskytovania služieb príjemcovi alebo jeho blokovaní či zrušení konta.¹³⁶ Online platformy musia disponovať vnútorným mechanizmom na vybavovanie sťažností, ktorý takýto príjemca môže využiť. DSA osobitne ustanovuje, že vybavenie sťažností nemôže byť založené výlučne na základe výstupov z automatizovaných prostriedkov.¹³⁷

V prípade, ak je sťažnosť vybavená, príjemcovia majú možnosť v prípade nespokojnosti s výsledkom využiť mechanizmus mimosúdneho riešenia sporov.¹³⁸

Online platformy majú možnosť pozastaviť poskytovanie svojich služieb na primeraný čas užívateľom, ktorí poskytujú zjavne nezákonný obsah. Taktiež majú online platformy právo dočasne pozastaviť vybavovanie oznámení od užívateľov, ktorí notifikačný systém zneužívajú a podávajú zjavne neopodstatnené oznámenia a sťažnosti.¹³⁹ Podmienkou pre takéto kroky je

¹³⁵ Bližšie pozri DSA, článok 15.

¹³⁶ DSA, článok 17 ods. 1.

¹³⁷ DSA, článok 17 ods. 5.

¹³⁸ DSA, článok 18.

¹³⁹ DSA, článok 20.

prvotné varovanie daných užívateľov. Kritériá pre vyhodnocovanie opodstatnenosti oznámení a sťažností musí platforma transparentne komunikovať vo svojich všeobecných obchodných podmienkach.

Veľmi veľké online platformy sú z hľadiska moderovania obsahu navyše povinné v zmysle článku 26 DSA **pravidelne posudzovať riziká** a prijímať opatrenia na ich zmierňovanie (článok 27). Posúdenie sú povinné vykonať raz ročne a je nutné zvýrazniť, že predmetná požiadavka sa netýka iba nezákonného obsahu, ale aj posúdenia:

“akýchkoľvek negatívnych účinkov na výkon základných práv na rešpektovanie súkromného a rodinného života, slobody prejavu a práva na informácie, zákazu diskriminácie a práv dieťaťa, ako sú zakotvené v článkoch 7, 11, 21 a 24 charty; a

úmyselnej manipulácie ich služieb, a to aj prostredníctvom neautentického používania alebo automatizovaného využívania služby, so skutočným alebo predvídateľným negatívnym vplyvom na ochranu verejného zdravia, maloletých osôb a občianskej diskusie alebo skutočnými alebo predvídateľnými účinkami súvisiacimi s volebnými procesmi a verejnou bezpečnosťou.”¹⁴⁰

Posudzovanie rizika preto nevyhnutne bude musieť zahŕňať aj šírenie dezinformácií na veľmi veľkých online platformách. Osobitne tomu nasvedčuje ustanovenie článku 26 ods. 2 DSA, ktoré vyžaduje špecificky skúmať **“ako systémy moderovania obsahu, odporúčacie systémy a systémy na výber a zobrazovanie reklamy** ovplyvňujú ktorékoľvek zo systémových rizík uvedených článku 26 ods. 1 vrátane možného rýchleho a rozsiahleho šírenia nezákonného obsahu a informácií, ktoré nie sú zlučiteľné s ich obchodnými podmienkami.”¹⁴¹

3.2.2.3 Transparentnosť, notifikačné a reportovacie povinnosti

DSA upravuje viaceré požiadavky týkajúce sa transparentnosti, či už v podobe podávania verejne dostupných správ, reklamy alebo odporúčacích systémov. Obsah a kvalita právnych požiadaviek na transparentnosť sa líši podľa toho, či ide o všeobecne sprostredkovateľov služieb, online platformy alebo veľmi veľké platformy. Otázky transparentnosti boli identifikované ako kľúčové odbornou komunitou pre získanie poznatkov o metódach a dôvodoch efektívneho šírenia dezinformácií.¹⁴²

Najvšeobecnejšia povinnosť transparentnosti sa nachádza v článku 12 DSA týkajúcom sa **všeobecných obchodných podmienok** sprostredkovateľských služieb. Tie okrem iného musia obsahovať aj informácie o **“akýchkoľvek politikách, postupoch, opatreniach a nástrojoch používaných na účely moderovania obsahu vrátane algoritmického rozhodovania a ľudskej kontroly.”**¹⁴³

¹⁴⁰ DSA, článok 26 ods. 1.

¹⁴¹ Zhodne SAVIN, A. The EU Digital Services Act: Towards a More Responsible Internet. Copenhagen Business School Law Research Paper Series No. 21-04. Dostupné na: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3786792.

¹⁴² Report of the independent High level Group on fake news and online disinformation. A Multi-Dimensional Approach to Disinformation, s. 22. Dostupné na: <https://op.europa.eu/sk/publication-detail/-/publication/6ef4df8b-4cea-11e8-be1d-01aa75ed71a1>.

¹⁴³ DSA, článok 12 ods. 1.

Pozornosť oprávnene vzbudzuje požiadavka pravidelne podávať **správy o transparentnosti**. Takúto správu musia zverejňovať sprostredkovateľské služby, ktoré nie sú mikropodniky alebo malé podniky, najmenej raz ročne.¹⁴⁴ Totožná povinnosť platí aj pre online platformy,¹⁴⁵ obsahovo však DSA pridáva viaceré náležitosti daných správ (pozri tabuľku č. 4). Veľmi veľké platformy taktiež musia zverejňovať správy o transparentnosti obohatené o informácie o vykonaných auditoch či analýzach rizík každých 6 mesiacov.¹⁴⁶

Tabuľka č. 4
Povinné náležitosti správ o transparentnosti podľa veľkosti platforiem
Zdroj: DSA.

Sprostredkovateľské služby (všeobecne) poskytnú informácie o	Online platformy poskytnú informácie o	Veľmi veľké online platformy poskytnú informácie o
počte príkazov prijatých od orgánov členských štátov rozdelených podľa druhu dotknutého nezákonného obsahu vrátane príkazov konať alebo poskytnúť informácie a o priemernom čase potrebnom na prijatie opatrení uvedených v týchto príkazoch	počte príkazov prijatých od orgánov členských štátov rozdelených podľa druhu dotknutého nezákonného obsahu vrátane príkazov konať alebo poskytnúť informácie a o priemernom čase potrebnom na prijatie opatrení uvedených v týchto príkazoch	počte príkazov prijatých od orgánov členských štátov rozdelených podľa druhu dotknutého nezákonného obsahu vrátane príkazov konať alebo poskytnúť informácie a o priemernom čase potrebnom na prijatie opatrení uvedených v týchto príkazoch
počte oznámení o nezákonnom obsahu rozdelených podľa druhu dotknutého údajne nezákonného obsahu, akýchkoľvek opatreniach prijatých na základe týchto oznámení rozlíšených podľa toho, či sa opatrenie prijalo na základe zákona alebo obchodných podmienok poskytovateľa, a o priemernom čase potrebnom na prijatie opatrení	počte oznámení o nezákonnom obsahu rozdelených podľa druhu dotknutého údajne nezákonného obsahu, akýchkoľvek opatreniach prijatých na základe týchto oznámení rozlíšených podľa toho, či sa opatrenie prijalo na základe zákona alebo obchodných podmienok poskytovateľa, a o priemernom čase potrebnom na prijatie opatrení	počte oznámení o nezákonnom obsahu rozdelených podľa druhu dotknutého údajne nezákonného obsahu, akýchkoľvek opatreniach prijatých na základe týchto oznámení rozlíšených podľa toho, či sa opatrenie prijalo na základe zákona alebo obchodných podmienok poskytovateľa, a o priemernom čase potrebnom na prijatie opatrení
moderovaní obsahu vykonávanom z vlastnej iniciatívy poskytovateľov vrátane počtu a druhu prijatých opatrení, ktoré ovplyvňujú dostupnosť, viditeľnosť a prístupnosť informácií poskytovaných príjemcami služby a ich schopnosť poskytovať informácie, rozdelených podľa druhu dôvodu a základu na prijatie týchto opatrení;	moderovaní obsahu vykonávanom z vlastnej iniciatívy poskytovateľov vrátane počtu a druhu prijatých opatrení, ktoré ovplyvňujú dostupnosť, viditeľnosť a prístupnosť informácií poskytovaných príjemcami služby a ich schopnosť poskytovať informácie, rozdelených podľa druhu dôvodu a základu na prijatie týchto opatrení;	moderovaní obsahu vykonávanom z vlastnej iniciatívy poskytovateľov vrátane počtu a druhu prijatých opatrení, ktoré ovplyvňujú dostupnosť, viditeľnosť a prístupnosť informácií poskytovaných príjemcami služby a ich schopnosť poskytovať informácie, rozdelených podľa druhu dôvodu a základu na prijatie týchto opatrení;
počte sťažností prijatých prostredníctvom vnútorného systému vybavovania sťažností, základe	počte sťažností prijatých prostredníctvom vnútorného systému vybavovania sťažností, základe	počte sťažností prijatých prostredníctvom vnútorného systému vybavovania sťažností, základe

¹⁴⁴ DSA, článok 13.

¹⁴⁵ DSA, článok 23.

¹⁴⁶ DSA, článok 33.

Sprostredkovateľské služby (všeobecne) poskytnú informácie o	Online platformy poskytnú informácie o	Veľmi veľké online platformy poskytnú informácie o
týchto sťažností, rozhodnutiach prijatých v súvislosti s týmito sťažnosťami, priemernom čase potrebnom na prijatie týchto rozhodnutí a počte prípadov, v ktorých boli tieto rozhodnutia zrušené	týchto sťažností, rozhodnutiach prijatých v súvislosti s týmito sťažnosťami, priemernom čase potrebnom na prijatie týchto rozhodnutí a počte prípadov, v ktorých boli tieto rozhodnutia zrušené	týchto sťažností, rozhodnutiach prijatých v súvislosti s týmito sťažnosťami, priemernom čase potrebnom na prijatie týchto rozhodnutí a počte prípadov, v ktorých boli tieto rozhodnutia zrušené
	počte sporov predložených orgánom mimosúdneho riešenia sporov, výsledky riešenia sporov a priemerný čas potrebný na dokončenie postupov riešenia sporov	počte sporov predložených orgánom mimosúdneho riešenia sporov, výsledky riešenia sporov a priemerný čas potrebný na dokončenie postupov riešenia sporov
	počte pozastavení, rozčlenených na pozastavenia uložené za poskytovanie zjavne nezákonného obsahu, predkladanie zjavne neopodstatnených oznámení a podávanie zjavne neopodstatnených sťažností	počte pozastavení, rozčlenených na pozastavenia uložené za poskytovanie zjavne nezákonného obsahu, predkladanie zjavne neopodstatnených oznámení a podávanie zjavne neopodstatnených sťažností
	akomkoľvek použití automatických prostriedkov na účely moderovania obsahu s uvedením presných účelov, ukazovateľov presnosti automatizovaných prostriedkov pri plnení týchto účelov a akýchkoľvek uplatnených záruk	akomkoľvek použití automatických prostriedkov na účely moderovania obsahu s uvedením presných účelov, ukazovateľov presnosti automatizovaných prostriedkov pri plnení týchto účelov a akýchkoľvek uplatnených záruk
		správu, v ktorej sa uvádzajú výsledky posúdenia rizika;
		súvisiace opatrenia na zmiernenie rizika
		audítorskú správu
		správu o vykonávaní odporúčaní auditu podľa článku

Ďalším inštitútom transparentnosti je **zverejňovanie informácií o reklame**, ktorá platí pre online platformy a s modifikáciami aj pre veľmi veľké platformy. Článok 24 DSA pre online platformy uvádza:

“Online platformy, ktoré vo svojich online rozhraniach zobrazujú reklamu, zabezpečia, aby príjemcovia služby mohli v prípade každej konkrétnej reklamy zobrazenej každému jednotlivému príjemcovi jasným a jednoznačným spôsobom a v reálnom čase identifikovať:

- a) že zobrazené informácie sú reklamou;
- b) fyzickú alebo právnickú osobu, v mene ktorej sa reklama zobrazuje;

- c) zmysluplné informácie o hlavných parametroch použitých na určenie príjemcu, ktorému sa reklama zobrazuje.”

Základnou myšlienkou danej požiadavky je, aby užívateľ pochopil, kto a prečo mu určitú reklamu ukazuje. Všeobecný zákaz personalizovanej reklamy, ako ho v reakcii na návrh DSA komunikoval Európsky parlament, sa zdá ako utópia.¹⁴⁷

Veľmi veľké online platformy sú navyše povinné prostredníctvom API sprístupniť a zverejniť archív reklamy, zobrazujúci špecifické informácie ukladané do jedného roka od posledného zobrazenia reklamy. Predmetný archív nesmie obsahovať žiadne osobné údaje príjemcov. Článok 30 ods. 2 ustanovuje:

“Archív musí obsahovať aspoň všetky tieto informácie:

- a) obsah reklamy;
- b) fyzickú alebo právnickú osobu, v mene ktorej sa reklama zobrazuje;
- c) obdobie, počas ktorého bola reklama zobrazená;
- d) informácie o tom, či sa reklama mala zobrazovať osobitne jednej alebo viacerým konkrétnym skupinám príjemcov služby, a ak áno, hlavné parametre použité na tento účel;
- e) celkový počet oslovených príjemcov služby a prípadne súhrnné počty za skupinu alebo skupiny príjemcov, ktorým bola reklama konkrétne určená.”

V neposlednom rade, DSA v kontexte veľmi veľkých online platforiem vyžaduje **transparentnosť odporúčacích systémov**. Právna požiadavka sa týka primárne sprístupnenia informácií o parametroch, ktoré odporúčacie systémy využívajú.

“Veľmi veľké online platformy, ktoré používajú odporúčacie systémy, vo svojich obchodných podmienkach **jasne, prístupným a ľahko zrozumiteľným spôsobom uvedú hlavné parametre** používané v ich odporúčacích systémoch, **ako aj všetky možnosti, ktoré príjemcovia služby majú na úpravu alebo ovplyvňovanie tých hlavných parametrov, ktoré im prípadne boli sprístupnené, vrátane aspoň jednej možnosti, ktorá nie je založená na profilovaní** v zmysle článku 4 ods. 4 nariadenia (EÚ) 2016/679 (GDPR).”¹⁴⁸

Ak veľmi veľká platforma poskytuje užívateľom viaceré možnosti (ne)využívania odporúčacích systémov, užívateľ musí mať v rozhraní platformy jednoduchú možnosť “kedykoľvek si vybrať a zmeniť uprednostňovanú možnosť pre každý z odporúčacích systémov určujúcu relatívne poradie informácií, ktoré sa mu predkladajú.”¹⁴⁹ Predpokladá sa, že uvedené požiadavky pre odporúčacie systémy ešte budú predmetom diskusií na úrovni orgánov EÚ a opozície platforiem z pohľadu ochrany práv duševného vlastníctva a obchodných tajomstiev.¹⁵⁰

¹⁴⁷ Európsky parlament. Report with recommendations to the Commission on a Digital Services Act: adapting commercial and civil law rules for commercial entities operating online. (2020/2019(INL)). Dostupné na: https://www.europarl.europa.eu/doceo/document/A-9-2020-0177_EN.html.

¹⁴⁸ DSA, článok 29 ods. 1.

¹⁴⁹ DSA, článok 29 ods. 2.

¹⁵⁰ BUITEN, M. The Digital Services Act: From Intermediary Liability to Platform Regulation. Working Paper. Dostupné na https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3876328, s. 22. Citované z BERBERICH, M. - SEIP, F. Der Entwurf des Digital Services Act. (2021) 1 Gewerblicher Rechtsschutz und Urheberrecht, Praxis im Immaterialgüter- und Wettbewerbsrecht (GRUR-Prax) 4.

3.2.2.4 Audity

Pre veľmi veľké online platformy DSA ustanovuje špecifické požiadavky na výkon externého auditu minimálne raz ročne. Cieľom auditu by malo byť posúdenie súladu s požiadavkami pre veľmi veľké online platformy predpísanými v DSA prípadne s kódexom správania, ktorý sa platforma zaviazala dodržiavať.¹⁵¹ Audity sú koncipované ako externé, nakoľko ich môžu vykonať iba nezávislé organizácie, ktoré majú dostatočne odborné kapacity a preukázali objektívnosť a etiku.¹⁵²

Výsledkom auditu by mala byť audítorská správa, ktorá okrem formálnych náležitostí musí obsahovať "audítorské stanovisko k tomu, či veľmi veľká online platforma, ktorá je predmetom auditu, dodržiavala povinnosti a záväzky v DSA, a to kladné, kladné s pripomienkami alebo záporné." V prípade negatívneho stanoviska správa musí zahŕňať aj "operatívne odporúčania týkajúce sa konkrétnych opatrení na dosiahnutie súladu" s DSA.¹⁵³

Veľmi veľké platformy sú povinné audítorskú správu zohľadniť a prijať opatrenia v nej navrhované. Do mesiaca od prijatia opatrení sú povinné prijať správu o plnení požiadaviek z audítorskej správy. "V prípadoch, keď operatívne odporúčania nevykonajú, uvedú v správe o vykonávaní odporúčaní auditu dôvody, prečo ich nevykonali, a stanovia alternatívne opatrenia, ktoré prípadne prijali na riešenie akýchkoľvek zistených prípadov nesúladu."¹⁵⁴

¹⁵¹ DSA, článok 28 ods. 1.

¹⁵² DSA, článok 28 ods. 2.

¹⁵³ DSA, článok 28 ods. 3.

¹⁵⁴ DSA, článok 28 ods. 4.

4. Diskusia a návrhy riešení

Časť 2 predkladanej štúdie prezentovala fungovanie online platforiem z hľadiska ekonomického modelu a technických možností optimalizácie zobrazovaného obsahu. Zo záverov tejto časti jednoznačne vyplýva potreba zásahu regulátora do fungovania online platforiem vzhľadom na ich ekonomickú silu, vplyv na spoločnosť a využívanie užívateľov ako produktu pre zvýšenie zisku. Zároveň sme načrtli a analyzovali využívanie odporúčacích systémov spolu s príkladmi ich využívania konkrétnymi sociálnymi médiami. Ako kľúčové problémy sme identifikovali transparentnosť generovaných výsledkov a nastavenie funkcie užitočnosti odporúčacích systémov v procese tréningu umelej inteligencie. Ako ďalšie problémy sme uviedli uzatváranie užívateľov do informačných bublín, skreslenia a férovosť pri odporúčaní a zbieranie spätnej väzby a odvodzovanie preferencií užívateľa v súvislosti s ochranou súkromia.

V časti 3 sme sa zamerali na analýzu právneho rámca predstaveného Európskou komisiou, ktorý sa venuje regulácii umelej inteligencie (AIA) a online platforiem (DSA). Zamerali sme sa špecificky na inštitúty, ktoré majú potenciál prispieť k efektívnejšiemu boju s dezinformáciami na internete v kontexte využívania umelej inteligencie.

V tejto časti štúdie prepojíme závery z časti 2 a zasadíme ich do právneho rámca prezentovaného v časti 3. Diskusiu priebežne obohacujeme o naše odporúčania pre reguláciu. Postupne prezentujeme naše myšlienky a návrhy vo všeobecnej rovine, k inštitútu auditu online platforiem, transparentnosti umelej inteligencie, správam o činnosti online platforiem a potrebe etického posúdenia rizík.

4.1 Všeobecné pripomienky k regulácii

4.1.1 Ekonomika pozornosti ako oblasť systémov AI vysokého rizika

Ako sme uviedli v časti týkajúcej sa analýzy AIA, využívanie umelej inteligencie online platformami momentálne vo všeobecnosti nie je vôbec navrhovanou legislatívou pokryté. V prípade zakázaných praktík máme za to, že využívanie systémov AI nebude možné v kontexte navrhovaných výnimiek zakázať, nakoľko ich výklad je stále veľmi nejasný a pri definovaní manipulatívnych praktík absentuje skúmanie podpory vlastných cieľov manipulujúcej osoby. Namiesto toho AIA na aplikáciu zákazu vyžaduje spôsobenie alebo možnosť spôsobenia ujmy jednotlivcom.

Navyše, oblasti pre systémy AI vysokého rizika nepokrývajú oblasť využívania umelej inteligencie online platformami. Ekonomický model online platforiem v podobe snahy o udržanie aktivity a pozornosti užívateľov síce môže byť legitímny, avšak v žiadnom prípade nie je bez rizika. Zásadným rizikom je aj vystavenie jednotlivca množstvu dezinformácií, ktoré môže mať za následok veľmi konkrétne rozhodnutia s negatívnymi dôsledkami na zdravie a život jednotlivca.

Vzhľadom na vyššie uvedené **odporúčame, aby AIA v prílohe III upravila samostatnú oblasť “ekonomiky pozornosti” alebo využívanie systémov AI za účelom získania pozornosti definovala ako aplikáciu systému AI vysokého rizika** v rámci 5. bodu prílohy III “Prístup k základným súkromným a verejným službám a dávkam a ich využívanie.”

Vítame kompromisné znenie AIA z dielne slovinského predsedníctva Rady EÚ, ktoré (čiastočne) rieši otázky uvedené vyššie. Všeobecný zákaz hodnotenia a klasifikácie fyzických osôb pre súkromné firmy môže zjemniť šírenie dezinformácií na internete. Ak však bude tento zákaz súčasťou regulácie, zároveň by mali aktivity spočívajúce v hodnotení a klasifikácii fyzických osôb na internete spadať pod systémy AI vysokého rizika z toho dôvodu, aby nezakázané praktiky nepôsobili v právnom vákuu.

4.1.2 Regulácia nezákonného a škodlivého obsahu

Jednou z najčastejších pripomienok k DSA v súvislosti so šírením dezinformácií je, že drvivá väčšina ustanovení DSA sa vzťahuje na nezákonný obsah. Je pritom zjavné, že dezinformácie nebudú vždy automaticky nezákonným obsahom. Výnimkou sú situácie, ak sa týkajú šírenia obsahu, ktorý je priamo právnym poriadkom zakázaný (napríklad extrémistické materiály, popieranie holokaustu) alebo nenávistných prejavov. Ostatná časť dezinformácií môže byť charakterizovaná “iba” ako škodlivý obsah. Rozumieme, že analýza škodlivosti obsahu môže byť v určitých prípadoch problematická a komplikovaná.

Súčasne navrhované znenie DSA ráta s analýzou rizikového škodlivého obsahu iba v súvislosti s požiadavkou posúdenia rizika pre veľmi veľké online platformy okrem iného aj hodnotením nastavenia odporúčacích systémov. Ako sme uviedli už v prvej časti, práve nastavenie funkcionality užitočnosti a zvýrazňovanie dezinformačného obsahu je jedným z kľúčových problémov. V prípade, ak veľmi veľká online platforma riziko šírenia dezinformácií identifikuje, mala by prijať opatrenia na zmiernenie či už prostredníctvom mechanizmov na moderovanie obsahu alebo odporúčacích systémov. Tieto platformy sú zároveň povinné zverejňovať správy o výsledkoch analýzy rizík verejnosti.

Pozícia pôsobnosti DSA voči nezákonnému a škodlivému obsahu nie je dokonca jednotná ani na úrovni Európskeho parlamentu. Ilustrovať to možno na nasledujúcich príkladoch:

- Výbor pre vnútorný trh a ochranu spotrebiteľa (IMCO) vo svojej správe obhajuje úzku pôsobnosť DSA na nezákonný obsah;
- Výbor pre občianske slobody, spravodlivosť a vnútorné veci (LIBE), naopak, navrhuje širší rozsah pôsobnosti aj na škodlivý obsah;
- Výbor pre právne veci (JURI) vo svojej správe vyjadruje nesúhlas s tým, aby online platformy určovali, čo je zákonné, a čo už nie.¹⁵⁵

Rozumieme, že priama kompetencia veľmi veľkých online platforiem posudzovať a odstraňovať škodlivý obsah je veľmi silná a môže byť ľahko zneužitelná. Takáto činnosť môže

¹⁵⁵ WINGFIELD, R. The Digital Services Act and Online Content Regulation: A Slippery Slope for Human Rights? Dostupné na <https://medium.com/global-network-initiative-collection/the-digital-services-act-and-online-content-regulation-a-slippery-slope-for-human-rights-eb3454e4285d>.

veľmi jednoducho limitovať výkon slobody prejavu a práva na informácie. Zároveň reflektujeme stav vedy a techniky, kde ani umelá inteligencia stále nevie dostatočne rozoznať škodlivý obsah.¹⁵⁶ Napriek tomu je príliš rizikové tváriť sa, že šírenie dezinformácií nie je škodlivé, nakoľko môže mať výrazný spoločenský dopad. Potreba istého legislatívneho zásahu je tak podľa nášho názoru odôvodnená.

Z týchto dôvodov **považujeme za nevyhnutné minimum, aby legislatíva obsahovala striktné pravidlá pre prípady, ak je zjavné, že online platforma či už priamo alebo nepriamo zvyrazňuje a uľahčuje šírenie dezinformácií.** Nakoľko nie je vhodné priamo platformám prikazovať vymazávanie škodlivého obsahu z hľadiska problematiky jeho identifikácie a potenciálnemu zásahu do slobody prejavu, považujeme za vhodné zakotviť mechanizmy, ktoré šírenie dezinformácií výrazne limitujú.

Za vhodné prostriedky považujeme:

- Požiadavky na transparentnosť a užívateľské voľby pri odporúčacích systémoch (viď časť 4.3. nižšie);
- Označovanie neovereného alebo neoveriteľného obsahu (content labelling);
- Zákaz propagovania určitého obsahu alebo tém (napríklad propagácia politických posolstiev alebo liečiv);
- Obmedzenie používania určitých metód pri citlivom obsahu (napríklad botov alebo mikro-targetingu pri politických reklamách).

Vítame výslovne uznanie šírenia dezinformácií ako systémového rizika v recitálvej časti kompromisného znenia DSA. Máme však za to, že sú potrebné špecifickejšie opatrenia uvedené vyššie.

4.2 Audity a nezávislá kontrola

4.2.1 Technologické aspekty vykonávania auditov

Ako sme ilustrovali v časti 2.3, zmeny v algoritmoch online platforiem (napr. v nastavení funkcie užitočnosti a jej váh) môžu často viesť aj k nezamýšľaným negatívnym dopadom na užívateľov. Ďalšie riziká v súvislosti s vývojom a nasadením umelej inteligencie sú v samotných tréningových dátach, ktoré môžu obsahovať rôzne skreslenia a môžu tak viesť k neférovému správaniu systémov AI k jednotlivcom alebo skupinám užívateľov, ale aj v spätnej väzbe, ktorú systém dostáva a používa ju na zlepšovanie budúcich predikcií.

Ide teda o komplexný problém, ktorý si vyžaduje komplexné riešenia, ktoré zahŕňajú **testovanie a auditovanie**. Testovanie je bežná praktika v softvérovom inžinierstve, ktorá sa realizuje pri vývoji softvéru na viacerých úrovniach, predovšetkým ako:

- *Jednotkové testovanie*, t. j. testovanie na úrovni jednotiek (napr. jednotlivých modulov) softvérového riešenia,

¹⁵⁶ MARSDEN, CH. - BROWN, I. - VEALE, M. Responding to Disinformation. Ten Recommendations for Regulatory Action and Forbearance. In Martin Moore & Damian Tambini (eds.) Dealing with Digital Dominance, OUP 2021, s. 205.

- *Integračné testovanie*, ktoré testuje interakciu medzi jednotlivými modulmi.

Problém je, že vývoj a nasadenie systémov AI zahŕňa tréningovanie modelov strojového učenia pomocou učiacich algoritmov, ktoré sú častou **čiernou skrinkou** (t. j. poznáme formát ich vstupu a výstupu aj vnútornú architektúru, ale ich správanie je, najmä v prípade hlbokých neuronových sietí, založené na komplexnej matematickej funkcii s miliónmi parametrov, ktoré častokrát nie je možné vyjadriť pomocou pre človeka jednoducho interpretovateľných pravidiel).

Zároveň takéto riešenia často trpia tzv. **technickým dlhom**, kedy jednotlivé fázy vývoja a prevádzky systémov AI nie sú dostatočne špecifikované, otestované a zdokumentované; v literatúre preto boli navrhnuté spôsoby hodnotenia pripravenosti systému AI pre produkčnú prevádzku.¹⁵⁷ Dôležité je tiež mať zadané jednoducho generovateľné testovacie príklady,¹⁵⁸ ktoré umožnia odhaliť možné problémy a skreslenia natrénovaných modelov nad rámec jedného, často skresľujúceho čísla o výkonnosti daného algoritmu (ako sú napr. presnosť, pokrytie a pod.; pozri časť 2.2.4).

Druhým prístupom riešenia problému negatívnych dopadov algoritmov online platforiem sú audity, ktoré tiež môžu byť realizované na viacerých úrovniach. Môžeme auditovať **procesy** online platforiem pri vývoji a monitorovaní algoritmov (napr. aká je úroveň ich technickej pripravenosti, aké sú procesy na identifikovanie negatívnych dopadov a ich riešenie a pod.), ale aj (a predovšetkým) **dáta použité na tréningovanie, algoritmy online platforiem a ich správanie v interakcii s užívateľmi**.

Zatiaľ čo to prvé sa najviac podobá na finančné audity, to druhé má svoje korene v auditovacích štúdiách zo sociálnych vied. V tradičnom ponímaní išlo o experimenty v teréne; prvé takéto experimenty boli realizované ekonómami z Ministerstva bývania a urbánneho rozvoja v USA (angl. US Department of Housing and Urban Development), aby identifikovali rasovú diskrimináciu v prístupe k bývaní.¹⁵⁹

V kontexte auditovania algoritmov a ich správania rozlišujeme viacero možných prístupov:^{160,161}

- **Audit zdrojového kódu** – základný typ auditu, kedy sa kontroluje zdrojový kód. Ako sme argumentovali vyššie, toto samotné nestačí, pretože správanie algoritmov na

¹⁵⁷ BRECK, E. – CAI, S. - NIELSEN, E. – SALIB M.- SCULLEY, D. The ML test score: A rubric for ML production readiness and technical debt reduction, 2017 IEEE International Conference on Big Data (Big Data), s. 1123-1132. Dostupné na: <https://doi.org/10.1109/BigData.2017.8258038>.

¹⁵⁸ Ako príklad uvádzame prístup navrhnutý pre testovanie správania modelov spracovania prirodzeného jazyka: RIBEIRO, M.T. - WU, T. - GUESTRIN, C. - SINGH, S. Beyond Accuracy: Behavioral Testing of NLP models with CheckList. Association for Computational Linguistics (ACL), 2020. Dostupné na: https://homes.cs.washington.edu/~marcotcr/acl20_checklist.pdf.

¹⁵⁹ SANDVIG, C. - HAMILTON, K. - KARAHALIOS, K. - LANGBORT, C. Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms. Data and Discrimination: Converting Critical Concerns into Productive Inquiry, a Preconference at the 64th Annual Meeting of the International Communication Association. May 22, 2014; Seattle, WA, USA., s. 1–23. Dostupné na: <https://social.cs.uiuc.edu/papers/pdfs/ICA2014-Sandvig.pdf>.

¹⁶⁰ Tamže.

¹⁶¹ BANDY, J. Problematic Machine Behavior: A Systematic Literature Review of Algorithm Audits. Proc. ACM Hum.-Comput. Interact. 5, CSCW1, Article 74 (April 2021), 34 pages. Dostupné na: <https://doi.org/10.1145/3449148>.

online platformách je dané natrénovanými modelmi. Zároveň je takýto typ auditu problematický z pohľadu práv duševného vlastníctva online platforiem. Nevyhnutným minimom je však transparentnosť na úrovni typu vstupných údajov, funkcie užitočnosti a pod. (pozri časť 4.3).

- **Neinvazívny audit užívateľov** – je založený na rozhovoroch s užívateľmi platformy o tom, aké sú ich skúsenosti a prípadné problémy pri jej používaní. Tento typ auditu však môže slúžiť len ako doplnkový, pretože nedokáže odhaliť príčiny pozorovaného správania platformy (kauzálne prepojenia).
- **Audit zahŕňajúci získavanie („scraping“) obsahu** – v tomto type auditu výskumník zadáva algoritmu platformy dopyty (napr. vyhľadávacie dopyty a pod.) buď pomocou poskytnutého programovacieho rozhrania (API) alebo priamym získaním („scrapovaním“) obsahu z webu platformy, čo však môže porušovať pravidlá používania online platforiem, pokiaľ sa to deje bez vedomia vlastníka platformy. Ďalšie obmedzenie predstavuje skutočnosť, že výskumník v tomto prípade nesimuluje správanie užívateľa, čo limituje možnosti auditovania personalizovaných systémov, akým sú odporúčacie systémy na online platformách.¹⁶² Na druhej strane umožňuje tento prístup nezávisle získať obsah, ktorý by inak poskytoval vlastník platformy netransparentným spôsobom.
- **Bábkový audit (angl. sock puppet audit)** – na rozdiel od predchádzajúceho typu je v tomto prípade simulované správanie užívateľov pomocou účelovo vytvorených profilov, ktoré vie výskumník ovládať (ako bábkoherec svoje bábkky, na základe čoho získal svoje meno). Veľkou výhodou tohto typu auditu je miera kontroly, ktorú výskumník (auditor) má nad (simulovaným) správaním užívateľov. To mu umožňuje skúmať, ako systém reaguje na rôzne typy spätnej väzby užívateľov, prípadne, ako sa správa pri rôznych typoch obsahu. Aj tu však môže byť problém s podmienkami používania online platforiem, ktoré účelové vytváranie profilov väčšinou zakazujú a viac či menej aktívne mu zamedzujú.
- **Kolaboratívny (crowdsourcingový) audit** – zahŕňa využitie reálnych užívateľov, ktorí súhlasia so zaznamenávaním svojho správania pre účely auditu a sú ochotní plniť zadané úlohy, ktoré sú predmetom auditu. Tento typ najvernejšie zachytáva správanie užívateľov. Na druhej strane si vyžaduje získanie dostatočne veľkej reprezentatívnej vzorky užívateľov a ich aktívne zapojenie počas vyžadovaného obdobia auditu (čo je v prípade dlhšie trvajúcich auditov problém). Keďže sú zapojení reálni užívatelia, sú s týmto auditom spojené aj väčšie riziká v súvislosti s osobnými údajmi účastníkov auditu. Všetky tieto problémy odpadajú, ak sa audit realizuje ako interný audit samotnou online platformou.

¹⁶² Vhodný je však napr. na auditovanie nepersonalizovaných algoritmov, napr. na rozpoznávanie a analýzu tvárí: RAJI, I.D. – BUOLAMWINI, J. Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES '19). Association for Computing Machinery, New York, NY, USA, s. 429–435. Dostupné na: <https://doi.org/10.1145/3306618.3314244>.

Navyše, audity môžu byť *interné* t.j. prevedené prevádzkovateľom algoritmu resp. audítorom majúcom prístup k algoritmu a dátam alebo *externé*, prevedené zvyčajne nezávislými výskumníkmi, ktorí skúmajú výstupy t.j. správanie auditovaného algoritmu.¹⁶³

V literatúre nájdeme viacero príkladov auditov na online platformách z posledného obdobia.¹⁶⁴ Z pohľadu možnosti kontroly podmienok auditu a jeho opakovateľnosti považujeme za najlepšie riešenie bábkový audit. Z technologického hľadiska je na realizáciu takéhoto auditu potrebné:

- **Naprogramovať botov (skripty), ktoré budú simulovať správanie užívateľa na platforme.** Interakcia môže byť jednoduchá (napr. sledovanie videí na platforme YouTube, ako sme realizovali v našej auditovacej štúdii¹⁶⁵) alebo komplikovanejšia, t. j., taká, ktorá bude vernejšie simulovať správanie užívateľa (napr. na platforme YouTube by to mohlo byť odovzdanie spätnej väzby na video v podobe „Páči sa mi to“, kliknutie na niektoré z odporúčaných videí, zapnutie sledovania kanálu, ktorý je autorom videa a pod.). Dá sa simulovať nový užívateľ na platforme (bez predchádzajúcej histórie) alebo užívateľ s istou históriou svojho správania na platforme. Podobne je možné porovnávať vplyv geografickej polohy, pohlavia a iných atribútov užívateľa na výsledné odporúčania.¹⁶⁶
- **Zaznamenávať obsah zobrazený (odporučený) platformou.** Dá sa zaznamenať odkaz na obsah (napr. video na YouTube) a naviazané metadáta (napr. počet pozitívnych hodnotení, počet komentárov, autor videa, nadpis a popis videa a pod.).
- **Vyhodnotiť zaznamenaný obsah.** V prípade dezinformácií hodnotenie spočíva v tom, či daná zaznamenaná položka obsahu podporuje niektorý dezinformačný naratív, alebo nie. Najčastejšie sa pritom používa hodnotenie pomocou zaškolených anotátorov, prípadne sa využíva múdrosť davu (crowdsourcing). Existujú aj snahy o automatickú anotáciu.¹⁶⁷

Takýto audit dokáže zistiť prevalenciu dezinformácií v odporúčaníach na online platforme a porovnať, ako sa mení v čase, resp. ktoré iné faktory majú na ňu vplyv. Porovnanie je však možné len vtedy, ak sa audit robí opakovane, čo je aktuálne pre jeho prínosnosť zriedkavé.

¹⁶² KNOTT, A. et al. Responsible AI for Social Media Governance: A Proposed Collaborative Method for Studying the Effects of Social Media Recommender Systems on Users. 2021. Available at: <https://gpai.ai/projects/responsible-ai/social-media-governance/responsible-ai-for-social-media-governance.pdf>.

¹⁶⁴ Pozri napr. kolaboratívny audit na Facebooku: SILVA, M. et al. Facebook Ads Monitor: An Independent Auditing System for Political Ads on Facebook. In Proc. of The Web Conference (WWW '20). ACM, New York, NY, USA, 2020, s. 224–234. Dostupné na: <https://doi.org/10.1145/3366423.3380109>. Prípadne bábkový audit na platforme YouTube: HUSSEIN, E. - JUNEJA, P. - MITRA, T. Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube. Proc. ACM Hum.-Comput. Interact. 4, CSCW1, Article 048, May 2020, 27 s. Dostupné na: <https://doi.org/10.1145/3392854>.

¹⁶⁵ TOMLEIN, M. - PECHER, B. - SIMKO, J. - SRBA, I. - MORO, R. - STEFANCOVA, E. - KOMPAN, M. - HRCKOVA, A. - PODROUZEK, J. - BIELIKOVA, M. An Audit of Misinformation Filter Bubbles on YouTube: Bubble Bursting and Recent Behavior Changes. RecSys '21: Fifteenth ACM Conference on Recommender Systems. Dostupné na: <https://doi.org/10.1145/3460231.3474241>.

¹⁶⁶ HUSSEIN, E. - JUNEJA, P. - MITRA, T. Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube. Proc. ACM Hum.-Comput. Interact. 4, CSCW1, Article 048, 2020. Dostupné na: <https://doi.org/10.1145/3392854>.

¹⁶⁷ PAPADAMOU, K. - ZANNETTOU, S. - BLACKBURN, J. - DE CRISTOFARO, E. - STRINGHINI G. - SIRIVIANOS. M. „It is just a flu“. Assessing the Effect of Watch History on YouTube's Pseudoscientific Video Recommendations, 2020. Dostupné na: <https://arxiv.org/abs/2010.11638>.

Prácnosť spočíva najmä v ručnom hodnotení (anotácii) zaznamenaného obsahu; pre ilustráciu, v nami realizovanej auditovacej štúdii citovanej vyššie sme nazbierali vyše 17 000 unikátnych videí, z ktorých sme ručne anotovali vyše 2900.

Ak teda chceme, aby bolo možné podobné audity vykonávať pravidelne (kontinuálne, napr. raz mesačne), je na to potrebné vyčleniť značné ľudské zdroje alebo využiť metódy na automatické anotovanie. To so sebou prináša niekoľko technických (a v súčasnosti aj výskumných) problémov, ako napr. zabezpečenie a kontrola férovosti natrénovaného modelu na automatickú anotáciu, zabezpečenie a kontrola toho, že metóda bude dobre zovšeobecňovať aj na nové dezinformačné naratívy a jej presnosť nebude v čase klesať a v neposlednom rade aj vernejšie simulovanie správania užívateľov a údržba skriptov na získavanie obsahu z online platformy. Výzvou je aj odlíšenie *exogénnych* vplyvov (zmeny v obsahu platformy spôsobené externými udalosťami, novým obsahom a pod.) od *endogénnych* (zmeny spôsobené zmenou algoritmov online platformy).¹⁶⁸

Technologicky sú teda audity algoritmov online platforiem a ich správania realizovateľné, aj keď je priestor na ich väčšiu automatizáciu a zníženie ich prácnosti. Zároveň sa ukazuje, že sú užitočné na analýzu správania online platforiem, odhaľovanie skreslení v algoritmoch, ale aj na nezávislú kontrolu skutočného stavu na online platformách a záväzkov, ku ktorým sa platformy samé zaviazali.

Problémom je aktuálne nejasný právny stav; výskumníci, ktorí auditovacie štúdie vykonávajú, často konajú nad rámec pravidiel používania online platforiem. Navyše, boli sme v nedávnej minulosti svedkami, kedy platformy potrestali výskumníkov, ktorí poukázali na problematiku správania platforiem, zablokovaním ich užívateľských profilov.¹⁶⁹

Sme za to, aby **sa zvýšila adopcia auditov na monitorovanie správania algoritmov online platforiem;** a to jednak interných auditov, ale keďže tieto majú svoje obmedzenia, ktoré uznávajú aj niektorí výskumníci z prostredia online platforiem¹⁷⁰, aj **nezávislých externých auditov**, vykonávaných auditormi podľa jasných pravidiel (pozri 4.2.2). Avšak, vzhľadom na výskumnú povahu viacerých problémov súvisiacich s algoritmi online platforiem, sa

¹⁶⁸ SIMKO, J. - TOMLEIN, M. - PECHER, B. - MORO, R. - SRBA, I. - STEFANCOVA, E. - HRCKOVA, A. - KOMPAN, M. - PODROUZEK, J. - BIELIKOVA, M.. Towards Continuous Automatic Audits of Social Media Adaptive Behavior and its Role in Misinformation Spreading. In Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '21). Association for Computing Machinery, New York, NY, USA, 2021, s. 411–414. Dostupné na: <https://doi.org/10.1145/3450614.3463353>.

¹⁶⁹ Išlo o výskumníkov z NYU, ktorí vykonávali kolaboratívny audit reklamy zobrazovanej užívateľom (pomocou nainštalovaného rozšírenia do webového prehliadača). Facebook ich účty zablokoval s odôvodnením, že získavali údaje mimo oficiálnych nástrojov platformy a zároveň, že kompromitovali aj súkromie užívateľov, ktorí s takým zaznamenávaním nesúhlasili. Dostupné na: <https://about.fb.com/news/2021/08/research-cannot-be-the-justification-for-compromising-peoples-privacy/>. Reakcia dotknutých výskumníkov je dostupná tu: <https://www.nytimes.com/2021/08/10/opinion/facebook-misinformation.html> a tu: <https://www.wired.com/story/facebooks-reason-banning-researchers-doesnt-hold-up/>. Ohradili sa v nej, že nezískavali žiadne osobné údaje, len údaje o reklamách a anonymizované údaje o zapojených užívateľoch (s ich súhlasom), pričom každý si môže skontrolovať fungovanie ich rozšírenia, pretože má otvorený zdrojový kód. Navyše, Facebook technicky ani legálne nezablokoval činnosť ich rozšírenia, ale zamedzil im prístup k dátam, ktoré sú prístupné pre iných užívateľov a výskumníkov.

¹⁷⁰ RAJI, I.D. - SMART, A. - WHITE, R.N. - MITCHELL, M. - GEBRU, T. - HUTCHINSON, B. - SMITH-LOUD, J. - THERON, D. - BARNES, P. Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20). Association for Computing Machinery, New York, NY, USA, s. 33–44. Dostupné na: <https://doi.org/10.1145/3351095.3372873>.

domnievame, že **existuje jednoznačný verejný záujem, aby nástroje auditovania mohli využívať aj výskumníci** za jasne stanovených podmienok (pozri časť 4.2.3).

4.2.2 Legislatíva a odporúčania

Vzhľadom na vyššie uvedené technologické aspekty vykonávania auditov a náš predchádzajúci výskum si dovoľujeme v prípade legislatívneho zakotvenia externých auditov uviesť niekoľko odporúčaní.

V prvom rade **považujeme za vhodné, aby Európska komisia upravila požiadavky na organizácie, ktoré audity v zmysle DSA môžu vykonávať**. Nastavené kritéria nezávislosti, odbornosti a etickosti výkonu auditov považujeme za príliš široké a vágne. Ako vhodným nastavením sa nám javí proces certifikácie audítorov prostredníctvom akreditovaných orgánov priamo Európskou komisiou. Podobný proces už dnes upravuje právo EÚ v oblasti kybernetickej bezpečnosti a ochrany osobných údajov. Zoznam certifikovaných audítorov by mal byť verejne dostupný.

V druhom rade **považujeme za nevyhnutné, aby priamo DSA legitimizovala zásahy do práv a záujmov online platforiem pri výkone auditov**. To znamená, že online platforma by audítorom nemohla odmietnuť prístup k určitým aktívam chránených právom na ochranu osobných údajov alebo právami duševného vlastníctva. Aby bol externý audit naozaj efektívny a plnil svoj účel, poverení audítori by mali mať čo najširší prístup k informáciám. Zároveň je vhodné, aby obchodné podmienky nezakazovali používanie botov výlučne za účelom vykonania bábkového auditu spomínaného vyššie.

Za ďalší nedostatok návrhu DSA považujeme **nedostatočnú záväznosť a vynútiteľnosť záveru auditov**. Veľmi veľké online platformy môžu odporúčania audítorov nahradiť vlastnými alternatívnymi opatreniami. Podľa nášho názoru takto formulovaná právna úprava umožňuje platformám nezohľadniť závery expertných audítorov a v konečnom dôsledku tieto závery ignorovať.

Zároveň **máme za to, aby DSA upravilo presný mechanizmus prieskumu implementácie záverov externých auditov** napríklad prostredníctvom automatickej kontroly koordinátora digitálnych služieb alebo príslušného orgánu či expertnej skupiny Európskeho výboru pre digitálne služby.

Požiadavky na výkon auditov sú v kompromisnom znení DSA ponechané až na detaily v pôvodnom znení. Víťame ale návrhy Európskeho parlamentu, ktorý navrhuje autorizovanie audítorov a široký prístup k dátam na účely výkonu auditov.

4.2.3 Pozícia vedeckého výskumu

Na tomto mieste považujeme za vhodné akcentovať pozíciu vedeckého výskumu v rámci vykonávania auditov online platforiem. DSA pozíciu akademického sektora zvyrazňuje prostredníctvom inštitútu prístupu k údajom, ktorý môžu využiť aj preverení vedeckí pracovníci

pridružení k akademickým inštitúciám. **Za problematické považujeme opätovne vážne kritéria pre určenie preverených výskumníkov a odporúčame mechanizmus podobný certifikácii so schvaľovacím procesom na úrovni EÚ.** Vo všeobecnosti oceňujeme ustanovenia týkajúce preverenia výskumníkov v kompromisnom znení DSA. Požiadavku na „nezávislosť od komerčných záujmov“ bez ďalšieho vysvetlenia však nevnímame ako šťastnú s ohľadom na výskumné inštitúcie financované zo súkromných zdrojov.

Zároveň **s obavami sledujeme snahy o zavedenie reštriktívnych opatrení pri prístupe k dátam pre vedecké inštitúcie** v podobe doplnenia ustanovení, ktoré by umožnili informácie nesprístupniť **z dôvodu ochrany práv duševného vlastníctva.** Máme za to, že ide o riešiteľnú situáciu prostredníctvom osobitných zmluvných dojednaní vrátane zmluvných pokút a záväzkov mlčanlivosti pre výskumníkov namiesto zákonného splnomocnenia pre zneprístupnenie údajov, ktoré pre výskum a porozumenie systémových rizík môžu byť dôležité. Ďalšie požiadavky pre výskumné účely môžu byť súčasťou kódexov správania.

4.3 Požiadavky na transparentnosť

Transparentnosť je jedným z kľúčových atribútov pre dôveryhodnú umelú inteligenciu.¹⁷¹ Predmetná hodnota môže mať v súvislosti so systémami AI viaceré podoby. Konkrétne povedomie verejnosti o využívaní umelej inteligencie (*awareness*), ukladanie záznamov a dohľadateľnosť procesov, ktoré viedli k určitému rozhodnutiu (*logging*) a samotnú vysvetliteľnosť konkrétnych rozhodnutí v konkrétnych prípadoch (*explainability*). V súvislosti so šírením dezinformácií sme už v časti 2 načrtli ako esenciálne problémy transparentnosť generovaných výsledkov a využívanie funkcie užitočnosti pri odporúčacích systémov. Máme za to, že všetky tri podoby transparentnosti uvedené vyššie, zohrávajú dôležitú úlohu v kontexte načrtnutých problémov.

Povedomie o využívaní umelej inteligencie sa z právneho hľadiska zvyšuje predovšetkým prostredníctvom rôznych informačných povinností. Priamo v súvislosti s umelou inteligenciou upravuje takúto povinnosť GDPR, kde prevádzkovatelia musia dotknuté osoby informovať o **existencii** „automatizovaného rozhodovania vrátane profilovania uvedeného v článku 22 ods. 1 a 4 GDPR a aspoň v týchto prípadoch zmysluplné informácie o **použitom postupe, ako aj význame a predpokladaných dôsledkoch** takéhoto spracúvania pre dotknutú osobu.“¹⁷² Problémom však je samotný článok 22 GDPR, na ktorý ustanovenie odkazuje. Na jeho aplikáciu sa totiž vyžaduje splnenie niekoľko veľmi nejasných požiadaviek (ako napríklad existencia právneho alebo podobného účinku pri automatizovanom individuálnom rozhodovaní) a samotná jeho podstata je stále diskutovaná.¹⁷³ Bežnou praxou preto je, že prevádzkovatelia vyhodnotia situáciu tak, že pri využívaní systémov AI sa na nich článok 22 GDPR nevzťahuje, a preto ani neinformujú o súvisiacich detailoch.

¹⁷¹ High-Level Expert Group on Artificial Intelligence (AI HLEG). Etické usmernenia pre dôveryhodnú umelú inteligenciu. Dostupné na: <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-sk/format-PDF>.

¹⁷² GDPR, článok 13 ods. 2 písm. f) a článok 14 ods. 2 písm. g).

¹⁷³ Napríklad TOSONI, L. The right to object to automated individual decisions: resolving the ambiguity of Article 22(1) of the General Data Protection Regulation. In International Data Privacy Law, Volume 11, Issue 2, April 2021, s. 145–162.

Z tohto dôvodu vítame informačné povinnosti týkajúce sa zobrazovania reklamy a transparentnosti odporúčacích systémov pre veľmi veľké online platformy prezentované v návrhu DSA. Je však nevyhnutné, aby informácie o použitých parametroch odporúčacích systémov boli poskytnuté zmysluplným spôsobom, a nie príliš všeobecne. Je absolútne nevyhnutné vyhodnotiť, ktoré parametre môžu spôsobovať efektívnejšie šírenie dezinformácií.

Zároveň považujeme za nevyhnutné, aby boli odporúčacie systémy pri prvom zapnutí online platformy štandardne (*by default*) vypnuté. Ak by užívateľ chcel túto funkciu využívať, musel by si ju zapnúť (*opt-in*). Prípadne by funkcia odporúčania bola rozdelená na menšie, reguláciou definované zložky (úrovne odporúčania), ktoré by si užívateľ podľa svojej voľby mohol zapnúť alebo vypnúť podľa toho, akú mieru odporúčania a personalizácie preferuje. Odporúčanie by sa napríklad mohlo skladať z dvoch úrovní: nepersonalizovanej a personalizovanej, pričom užívateľ by si mohol zapnúť napr. len tú nepersonalizovanú.

Máme za to, že toto pomerne jednoduché technické riešenie dá do rúk užívateľom väčšiu autonómiu od prvého kontaktu s online platformou a môže znížiť zobrazovanie škodlivého obsahu. Toto pravidlo by však malo platiť takým spôsobom, aby od účinnosti DSA boli platformy povinné odporúčacie systémy vypnúť aj existujúcim užívateľom. Sme si vedomí navrhovanej zmeny daných ustanovení v kompromisnom znení DSA. Nie sme však presvedčení, že navrhovaná formulácia v podobe požiadavky odporúčacích systémov „nenarúšať autonómiu, rozhodovanie alebo voľbu príjemcu služby prostredníctvom dizajnu, štruktúry, funkcie alebo spôsobu fungovania ich online rozhrania“ bude dostatočná vzhľadom na pripomienky uvedené vyššie.

Uchovávanie logov bude v rámci inštitútov správy údajov vyžadovať AIA. Práve dosledovateľnosť rozhodnutí systémov AI vysokého rizika je explicitne v návrhu AIA uvedená ako dôvod pre vedenie záznamov (logov). Aj z tohto dôvodu považujeme za nevyhnutné, aby AIA klasifikovala ekonomiku pozornosti ako oblasť vysokého rizika.

Ak by ostal v tejto veci zachovaný pôvodný návrh, dosledovateľnosť rozhodnutí systémov AI v súkromnom sektore by v zásade nemala právnu oporu. V takom prípade odporúčame, aby finálne znenie DSA obsahovalo povinnosť uchovávať logy týkajúce sa odporúčacích systémov.

Máme za to, že právo na vysvetlenie konkrétnych rozhodnutí systémami AI v súčasnosti v právnom poriadku neexistuje.¹⁷⁴ Podľa nášho názoru by súčasný stav vedy a techniky vo veľmi veľa prípadoch nedokázal takúto právnu povinnosť pokryť. **Mali by sme ale smerovať k takému stavu, ktorý súčasný stav techniky pokryť dokáže, a to minimálne prostredníctvom informácií o spracovaných dátach a využitých parametroch a ich hodnotách.** Tieto informácie môžu byť veľmi cenné osobitne pri boji s dezinformáciami a nastavení mechanizmov na zníženie ich šírenia. Považujeme za vhodné, aby užívatelia aspoň v minimálnej miere vedeli,

¹⁷⁴ WACHTER, S. - BITTELSTATD, B. - FLORIDI, L. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. International Data Privacy Law, Volume 7, Issue 2, May 2017, s. 76–99.

prečo je im špecifický obsah zobrazovaný, a mali možnosť parametre zobrazovania obsahu modelovať.

4.4 Správy o činnosti platforiem

Správy o činnosti platforiem odzrkadľujú princíp transparentnosti a právo na informácie verejnosti o činnosti online platforiem. Malo by ísť ale o informácie relevantné a presne určené návrhom DSA. Okrem toho DSA predpokladá, že sa veľmi veľké platformy zaviazu plniť kódexy správania; v kontexte dezinformácií pôjde o kódex, ktorý vznikne posilnením Kódexu nakladania s dezinformáciami. Aj z tohto kódexu vyplývajú záväzky ohľadom reportovania aktivít veľmi veľkých platforiem v súvislosti s dezinformáciami. Usmernenie Európskej komisie k posilneniu kódexu pritom zdôrazňuje potrebu toho, aby signatári kódexu „poskytovali informácie a údaje na monitorovanie v štandardizovaných formátoch, rozčlenené podľa členských štátov a včas.“¹⁷⁵ **Podporujeme reportovanie rozčlenené podľa členských štátov.** Podľa nášho názoru by sa rovnaký princíp mal zaviesť aj pri poskytovaní informácií v rámci správ o činnosti platforiem, ktoré ustanovuje DSA. Víťame návrhy Európskeho parlamentu, ktoré túto požiadavku vo svojich návrhoch reflektujú.

Zároveň, máme za to, že drvivá väčšina požiadaviek pre správy o činnosti platforiem sa týka štatistických ukazovateľov. Podľa nášho názoru **by bolo prínosné, aby online platformy v rámci správ informovali aj o konkrétnom riešení konkrétnych prípadov.** Ako ilustráciu je možné uviesť správy o stave ochrany osobných údajov, ktoré každoročne publikuje Úrad na ochranu osobných údajov SR.¹⁷⁶ V daných správach sa nachádzajú konkrétne prípady a ich analýza z hľadiska skutkového a právneho stavu. Pre prevádzkovateľov osobných údajov tak predstavujú pomerne presné návody na riešenie určitých situácií. Zároveň má verejnosť informácie o postupe dozorných orgánov v niektorých prípadoch.

Podľa nášho názoru by zverejňovanie riešení konkrétnych prípadov online platformami ešte zvýraznilo požiadavky transparentnosti a zároveň prezentovalo silné a slabé stránky moderovania obsahu na sociálnych sieťach.

4.5 Posudzovanie z pohľadu etiky

V predchádzajúcich častiach sme poukázali na to, prečo by fungovanie online platforiem a činnosť ich algoritmov mali v určitej miere podliehať zásahu zo strany regulátora. Táto regulácia by však nemala prichádzať len zo strany štátov alebo medzinárodných organizácií, ale aj zo strany samotných online platforiem, pričom jej cieľom by mala byť proaktívna identifikácia možných rizík s ohľadom na kľúčové spoločenské hodnoty a etické princípy.

¹⁷⁵ Európska komisia. OZNÁMENIE KOMISIE EURÓPSKEMU PARLAMENTU, RADE, EURÓPSKEMU HOSPODÁRSKEMU A SOCIÁLNEMU VÝBORU A VÝBORU REGIÓNOV Usmernenie Európskej komisie k posilneniu Kódexu postupov proti šíreniu dezinformácií. COM/2021/262 final.

Dostupné na: <https://eur-lex.europa.eu/legalcontent/SK/TXT/?uri=CELEX%3A52021DC0262#>.

¹⁷⁶ Pozri napríklad Úrad na ochranu osobných údajov Slovenskej republiky. Správa o stave ochrany osobných údajov za rok 2020. Dostupné na: <https://dataprotection.gov.sk/uouu/sk/content/vyrocne-spravy>.

Pri posudzovaní dopadov činnosti online platforiem je nanajvýš dôležité identifikovať rôzne dotknuté skupiny osôb či organizácií, ktoré platforma svojou činnosťou môže zasiahnuť. Tými nemusia byť len samotní užívatelia, resp. tzv. priami zainteresovaní (angl. stakeholders), ktorí prichádzajú do bezprostredného kontaktu s platformou. Rovnako nimi môžu byť aj ďalšie, častokrát veľmi zraniteľné skupiny, ktoré si vyžadujú našu pozornosť.¹⁷⁷ Treba zdôrazniť, že každá z týchto skupín si zasluhuje pozornosť prevádzkovateľov online platforiem bez ohľadu na to, či predstavuje priameho užívateľa platformy, resp. generuje profit pre túto platformu.

Správna a včasná identifikácia dotknutých osôb, ako aj možných spoločenských a etických rizík nie je triviálna a je aj predmetom takzvanej Colingridgovej dilemy.¹⁷⁸ Teda situácie, kde na začiatku vývoja nejakej technológie, kedy máme ešte veľkú mieru kontroly nad jej finálnou podobou, je naše poznanie o jej možných dopadoch na ľudí značne obmedzené, avšak po jej nasadení do života sú zas obmedzené naše schopnosti túto technológiu kontrolovať.

Dotknuté osoby, rozsah rizík, ale aj miera ich dopadu sa navyše môžu v čase meniť, či už pod vplyvom technologického rozvoja, alebo stavu nášho poznania o týchto rizikách.

Je preto dôležité, aby platformy dokázali vyčleniť **čas a zdroje na kontinuálne vyhodnocovanie** možných dopadov z pohľadu **dotknutých morálnych hodnôt a princípov počas celého cyklu** od dizajnu nových funkcionalít až po ich nasadenie.

Vyvstáva otázka, či je vôbec možné definovať sadu nejakých univerzálnych princípov a hodnôt, ktoré by sme mali ako spoločnosť nasledovať.¹⁷⁹ Ako ukázal nedávny výskum, ktorý analyzoval desiatky významných iniciatív na poli etickej a dôveryhodnej umelej inteligencie¹⁸⁰, dokážeme v ich textoch identifikovať sadu istých univerzálnych hodnôt alebo okruhov princípov, ku ktorým sa všetky tieto iniciatívy opakovane obracajú, a na ktoré kladú primárny dôraz. Medzi tieto hodnoty patria napr. už spomínaná transparentnosť, ale aj férovosť či ľudské súkromie.

Podobné okruhy nájdeme napríklad aj medzi požiadavkami na dôveryhodnú umelú inteligenciu tak, ako ich sformulovala Expertná skupina na vysokej úrovni pre umelú inteligenciu (AI HLEG) v už spomínaných Etických usmerneniach pre dôveryhodnú umelú inteligenciu (pozri obrázok č. 10).

Aktuálne sa v rámci AI etiky prešlo fázou ujasňovania si základných princípov a hodnôt. Pozornosť expertov v tejto oblasti sa preto presúva do tzv. operacionalizácie, teda ich prekladu do konkrétnych krokov a požiadaviek, ktoré je potrebné v prípade návrhu, vývoja, nasadzovania a používania systémov AI dodržiavať.

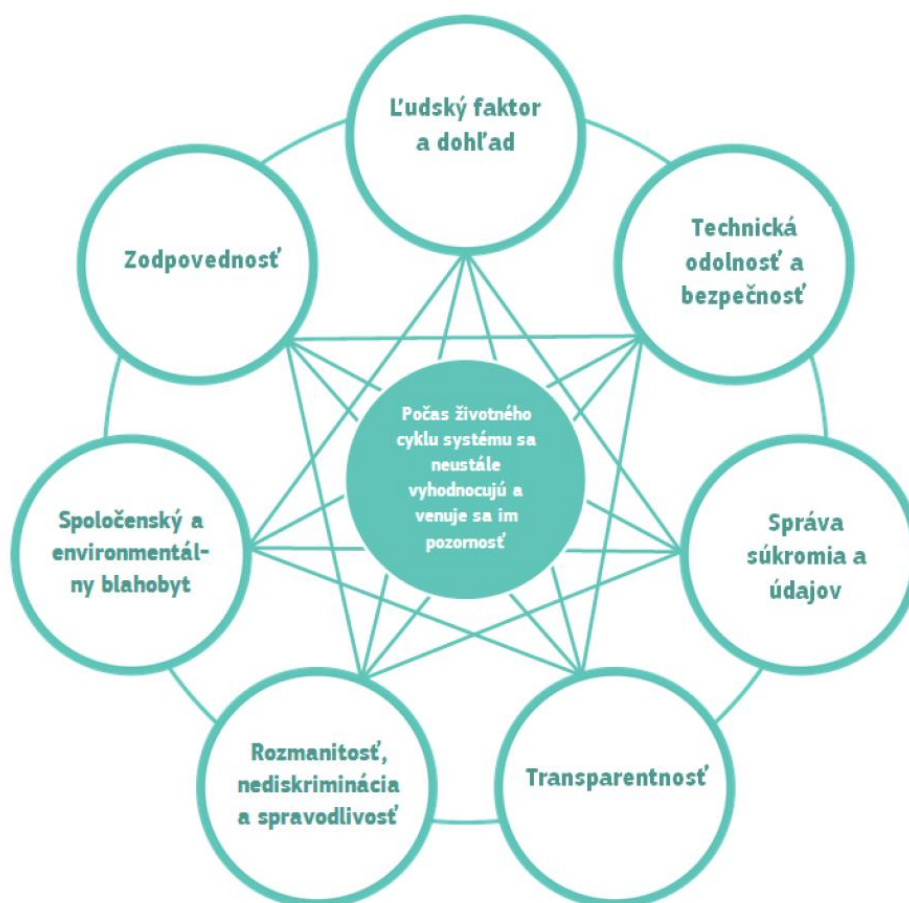
¹⁷⁷ FRIEDMAN, B. - KAHN J, P. – BORNING, A. Value Sensitive Design and Information Systems. In: DOOR, N. et al. (eds). Early engagement and new technologies: Opening up the laboratory. Philosophy of Engineering and Technology, vol 16. Springer, Dordrecht. Dostupné na: https://doi.org/10.1007/978-94-007-7844-3_4.

¹⁷⁸ COLLINGRIDGE, D. The Social Control of Technology. New York: St. Martin's Press, 1980.

¹⁷⁹ Ide o jednu z významných debát, ktoré sa odohrávajú hlboko v srdci filozofického skúmania a tvoria časť tzv. metatetiky. Aj keď existuje niekoľko silných argumentov v prospech etického relativizmu, v súčasnosti skôr prevláda medzi odborníkmi názor, že takéto univerzálne morálne princípy a hodnoty vieme stanoviť. Pozri napr. RACHELS, S. – RACHELS, J. The Elements of Moral Philosophy. McGraw-Hill Education, 2015.

¹⁸⁰ FIELD, J. et al. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI. Berkman Klein Center for Internet & Society, 2020.

V súčasnosti existuje viacero rôznych nástrojov pre posúdenie systémov umelej inteligencie z pohľadu AI etiky.¹⁸¹ Medzi tie najkomplexnejšie patrí napríklad **Zoznam bodov na posudzovanie dôveryhodnej umelej inteligencie** (Assessment List for Trustworthy AI - ALTAI), ktorý je výsledkom činnosti práve expertnej skupiny AI HLEG. Tento zoznam, ako aj niektoré iné nástroje na posúdenie spoločenských dopadov AI,¹⁸² sa už snažia aj o spomínanú operacionalizáciu hodnôt a princípov, ktoré si ctíme v našej spoločnosti a ktorých dodržiavanie považujeme za kľúčové pre ďalšie fungovanie demokratickej spoločnosti.



Obrázok č. 10
Požiadavky dôveryhodnej umelej inteligencie
Zdroj: Ethics Guidelines for Trustworthy AI (AI HLEG).

Prítomnosť etických úvah v umelej inteligencii a snaha o samoreguláciu jej systémov môžu byť veľmi nápomocné aj s ohľadom na situáciu, v ktorej sa pri regulácii nových technológií nachádzame. Aktuálne sme totiž pri regulácii umelej inteligencie v stave, ktorý britský filozof

¹⁸¹ AYLING, J. – CHAPMAN, A. Putting AI ethics to work: are the tools fit for purpose?. AI Ethics, 2021. Dostupné na: <https://doi.org/10.1007/s43681-021-00084-x> ponúka veľmi pekný prehľad desiatok nástrojov pre potreby vyhodnotenia z pohľadu AI etiky. Okrem toho, že zhrňa a kategorizuje najvýznamnejšie nástroje pre posudzovanie systémov AI, pridáva aj historický kontext pre tieto nástroje z iných oblastí (enviro, finančný sektor a pod.)

¹⁸² Zmieniť treba aj veľmi komplexné postupy z dielne IEEE, ktoré sa venujú posudzovaniu dopadov autonómnych a inteligentných systémov na ľudské dobro, napr. Štandard IEEE 7010-2020. Dostupné na: <https://standards.ieee.org/standard/7010-2020.html> alebo posúdenie vplyvov z pohľadu dátovej etiky určené pre verejné inštitúcie v UK. Dostupné na: <https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020>.

James Moor pomenoval ako tzv. **vákuum politik**¹⁸³. Ide o situáciu, keď sa politiky pre reguláciu digitálnej technológie ešte len tvoria a častokrát ani samotné kľúčové pojmy, ktoré sú pre diskurz potrebné, nie sú ešte ani zďaleka ustálené. V tejto nestabilnej situácii sa preto ozýva volanie po menšej miere tvrdej regulácie umelej inteligencie a silnejšej samoregulácii, ktorá by pomohla preklenúť tieto medziobdobia neistoty.

Túto príležitosť však regulátori nevyužívajú v dostatočnej miere. Ako sme aj uviedli vo svojej pozícii voči návrhu Aktu o umelej inteligencii (AIA),¹⁸⁴ v texte návrhu sa uvádzajú etické dopady len v nezáväznej forme odôvodnení a v dôvodovej správe. Podľa nás to predstavuje premárnenú príležitosť s ohľadom na prácu, ktorú v tomto smere vykonal AI HLEG. **Posúdenie etických rizík by sa podľa nás malo považovať za záväznú súčasť posudzovania zhody navrhovaného pre poskytovateľov systémov AI v AIA.** Dobré zvolené nástroje na posudzovanie systémov AI z pohľadu etiky, ktoré už reálne fungujú v praxi,¹⁸⁵ tak môžu napomôcť aj pri preklenutí už spomínaného vákua politik a obdobia, kedy navrhované regulácie ešte nie sú naplno účinné.

Na druhej strane, jedným z rizík prílišného spoliehania sa na samoreguláciu je, že sa bude vykonávať len formálne, čo vidíme aj v prípade interných auditov online platforiem a správ o ich činnosti. V prípade takto predstieranej samoregulácie digitálnych technológií sa používa termín *ethics bluwashing*.¹⁸⁶ A teda, podobne ako v prípade greenwashingu, keď nejaká spoločnosť len predstiera, že koná v záujme ochrany životného prostredia, tak aj tu sa môže stať, že sa etickej a spoločenskej reflexii budú spoločnosti venovať len naoko, a aby sa vyhlili priamej regulácii zo strany štátov.

Ako ideálny sa potom javí model tzv. **hybridnej regulácie**, kedy súčasne pôsobia na subjekty a dotknuté osoby viaceré regulačné nástroje súčasne.¹⁸⁷

4.6 Závery

Ak vezmeme do úvahy čiastkové závery prezentované v časti 2 predkladanej štúdie a špecifická skúmaných návrhov legislatívy z dielne Európskej komisie (hlavne návrh nariadenia o umelej inteligencii a návrh nariadenia o digitálnych službách) prezentované v časti 3, máme za to, že skúmané návrhy právnych aktov je možné vylepšiť takým spôsobom, aby poskytovali účinnejšie nástroje pre boj so šírením dezinformácií.

¹⁸³ MOOR, J. What Is Computer Ethics? Metaphilosophy 16(4), 1985, s. 266 – 275.

¹⁸⁴ Kempelen Institute of Intelligent Technologies. Stance on the Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. Dostupné na: https://kinit.sk/wp-content/uploads/2021/09/KINIT_Stance-of-AIA_Paper_2021_09.pdf.

¹⁸⁵ Interaktívna podoba nástroja ALTAI je k dispozícii verejne a zdarma, pričom už teraz poskytuje aj pomerne slušný užívateľský komfort <https://altai.insight-centre.org/>.

¹⁸⁶ FLORIDI, L. Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. Philos. Technol. 32, 2019, s. 185–193.

¹⁸⁷ Môžeme spomenúť napríklad štyri modalita regulácie spoločenských vzťahov od L. Lessiga, kde všetky modalita sa navzájom dopĺňajú a fungujú popri sebe (právo, etika, trh a architektúra/kód). LESSIG, L. Laws of Cyberspace. Dostupné na: https://cyber.harvard.edu/works/lessig/laws_cyberspace.pdf.

Z hľadiska všeobecných pripomienok **považujeme za kľúčové, aby oblasť ekonomiky pozornosti bola považovaná za oblasť systémov AI vysokého rizika**, aby využívanie algoritmov sociálnymi médiami „neušlo“ legislatívnym požiadavkám prezentovaných v AIA.

Zároveň rozumieme limitom právnej úpravy škodlivého obsahu z hľadiska zásahov do slobody prejavu a priznaniu príliš veľkej moci online platformám. Na druhej strane, legislatíva môže poskytovať široké penzum nástrojov, ktoré majú potenciál limitovať šírenie dezinformácií na online platformách bez priamej regulácie škodlivého obsahu.

Ako príklady zavedenia takýchto „nepriamych“ pravidiel odporúčame:

- Požiadavky na transparentnosť a užívateľské voľby pri odporúčacích systémoch;
- Označovanie neovereného alebo neoveriteľného obsahu (*content labelling*);
- Zákaz propagovania určitého obsahu alebo tém;
- Obmedzenie používania určitých metód pri citlivom obsahu (napríklad botov alebo mikro-targetingu pri politických reklamách).

Za absolútne zásadné považujeme výkon externých auditov. DSA síce takýto mechanizmus priamo upravuje, ale s obavami sledujeme jeho limity týkajúce sa nedostatku jasných pravidiel pre určenie vhodných entít pre vykonanie externých auditov, nedostatočnej záväznosti výsledkov takýchto auditov a implementáciu ich záverov. Zároveň, prístup externých audítorov by nemal byť obmedzovaný z hľadiska ochrany práv online platformy ako obchodné tajomstvo alebo tretích strán v podobe ochrany osobných údajov. Z technického hľadiska odporúčame ako vhodnú formu auditu tzv. bábkový audit, avšak vyššie uvedené pripomienky je nutné nastaviť takým spôsobom, aby využívanie botov na účely externého auditu bolo z hľadiska legislatívy a podmienok využívania online platforiem možné. Zvýrazňujeme taktiež rolu vedeckého výskumu a prístup k údajom zo strany preverených výskumníkov.

Dôležitou oblasťou pri boji s dezinformáciami je aj transparentnosť. V tomto smere vítame návrhy prezentované v DSA týkajúce sa transparentnosti reklamy a odporúčacích systémov či verejne dostupných správ. Tieto mechanizmy ale musia zabezpečovať, aby ich výsledky boli zmysluplné a ukazovali odbornej a laickej verejnosti skutočný stav. **V súvislosti s transparentnosťou odporúčacích systémov upriamujeme pozornosť na to, aby DSA zakotvila v akejkoľvek forme povinný opt-in pre užívateľov pri prvom kontakte s online platformou.** Odporúčanie by sa napríklad mohlo skladať z dvoch úrovní: nepersonalizovanej a personalizovanej, pričom užívateľ by si mohol zapnúť napr. len tú nepersonalizovanú, ktorá pri odporúčaní neberie do úvahy správanie užívateľa, ktoré je zdrojom pre odhad preferencií. Zároveň by uchovávanie logov pre odporúčacie systémy malo byť právne zakotvené v DSA pre prípady, ktoré by potenciálne AIA nepokrývala.

Správy o transparentnosti by mali byť doplnené o riešenie konkrétnych prípadov, aby bolo jasné, ako sa online platforma v konkrétnych situáciách správa a aké opatrenia na zmierňovanie dopadov prijíma. Zároveň by štatistické ukazovatele v týchto správach mali byť rozčlenené podľa členských štátov, aby bolo možné odhaliť prípadné rozdiely v šírení

dezinformačného obsahu v jednotlivých členských štátoch, ako aj v prístupe platforiem a orgánov členských štátov a v prijatých opatreniach.

Nemenej dôležité je, aby sociálne médiá pravidelne vyhodnocovali nielen právne, ale aj etické riziká. Považujeme za zásadné, aby platformy dokázali vyčleniť čas a zdroje na kontinuálne vyhodnocovanie možných dopadov z pohľadu dotknutých morálnych hodnôt a princípov počas celého cyklu od dizajnu nových funkcionalít až po ich nasadenie.

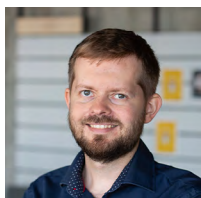
Posúdenie etických rizík by sa podľa nás malo považovať za záväznú súčasť posudzovania zhody navrhovaného pre poskytovateľov systémov AI v AIA.

Autori



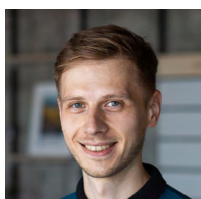
Matúš Mesarčík

Špecialista na etiku a právo
kontakt: matus.mesarcik@kinit.sk



Róbert Móro

Výskumník v oblasti umelej inteligencie a strojového učenia
kontakt: robert.moro@kinit.sk



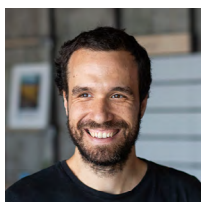
Michal Kompan

Výskumník v oblasti umelej inteligencie a strojového učenia
kontakt: michal.kompan@kinit.sk



Juraj Podroužek

Výskumník v oblasti etiky digitálnych technológií
kontakt: juraj.podrouzek@kinit.sk



Jakub Šimko

Výskumník v oblasti umelej inteligencie a strojového učenia
kontakt: jakub.simko@kinit.sk



Mária Bieliková

Výskumníčka v oblasti umelej inteligencie a strojového učenia
kontakt: maria.bielikova@kinit.sk