STŘEDOŠKOLSKÁ ODBORNÁ ČINNOST

Obor: 18. Informatika

Auditování algoritmů sociálních sítí s využitím umělé inteligence

Social media AI Algorithm Auditing

Autor: Erik Macák, Martin Pytlík, Michal Tvaroh
Škola: DELTA – Střední škola informatiky a ekonomie, s.r.o.,
Ke Kamenci 151, 530 03, Pardubice
Kraj: Pardubický
Konzultant: Ivan Srba, PhD
Rok: 2025/2026

Prohlášení

Prohlašuji, že jsme svou práci SOČ vypracovali samostatně a použili jsme pouze prameny a literaturu uvedené v seznamu bibliografických záznamů.

Beru na vědomí, že nejpozději odevzdáním slovesné vědecké práce do veřejné soutěže Středoškolská odborná činnost, stejně jako odevzdáním jejích příloh a dalších připojených děl, např. audiovizuálních, fotografických, výtvarných, architektonických apod. (dále jen „soutěžní dílo“), dochází ke zveřejnění díla podle § 4 odst. 1 zákona č. 121/2000 Sb., autorského zákona, ve znění pozdějších předpisů (dále jen „autorský zákon“). Totéž platí pro pozdější odevzdání doplněného, změněného, upraveného nebo opraveného díla.

Beru na vědomí, že zveřejněním díla, jehož součástí je vynález, se tento vynález stává součástí stavu techniky podle § 5 odst. 1, 2 zákona č. 527/1990 Sb., o vynálezech, průmyslových vzorech a zlepšovacích návrzích, ve znění pozdějších předpisů (dále jen „patentový zákon“), což zakládá překážku pro udělení patentu podle § 3 odst. 1 patentového zákona.

Beru na vědomí, že vyhlašovatel soutěže je podle § 61 odst. 1 autorského zákona per analogiam oprávněn užít soutěžní dílo pro účely zajištění průběhu soutěže, zejména k zajištění transparentnosti soutěže a veřejnosti obhajob soutěžních prací. V odůvodněném rozsahu je tedy vyhlašovatel po dobu účasti autora v soutěži oprávněn zejména:

- zhotovovat rozmnoženiny díla, je-li to nezbytné k seznámení účastníků soutěže, porotců nebo veřejnosti se soutěžní prací;
- zapůjčit originál nebo rozmnoženinu díla účastníkům soutěže, porotcům nebo veřejnosti. Přitom dbá na bezpečné nakládání s dílem;
- vystavovat originál nebo rozmnoženinu díla v průběhu soutěžních přehlídek a doprovodných akcí;
- sdělovat dílo veřejnosti v nehmotné podobě, a to především počítačovou nebo obdobnou sítí.

Dále prohlašuji, že při tvorbě této práce jsme použili nástroj generativního modelu AI [ChatGPT 5.3; https://chatgpt.com/] za účelem pomoci se strukturováním textu. Po použití tohoto nástroje jsme provedli kontrolu obsahu a přebírám za něj plnou zodpovědnost.

V Pardubicích dne 24.3.2026 Erik Macák

V Pardubicích dne 24.3.2026 Martin Pytlík

V Pardubicích dne 24.3.2026 Michal Tvaroh

Poděkování

Mnohokrát děkujeme celému institutu KInIT za představení problému a možnost spolupráce. Speciální poděkování patří panu doktoru Ivanu Srbovi za vedení práce a dohled na morálku celého týmu.

Anotace

Práce se zabývá návrhem a realizací systému pro automatizované algoritmické audity doporučovacích feedů sociálních sítí. Teoretická část popisuje problémy spojené s doporučovacími algoritmy – dezinformace, psychologické dopady, selhání moderace obsahu – a zasazuje je do kontextu evropské regulace DSA a jejích nedostatků při vymáhání transparentnosti platforem. Praktická část představuje třísložkový systém: mobilního agenta ovládajícího aplikaci Instagram Reels prostřednictvím knihovny UIAutomator2, prediktivního systému simulujícího chování syntetických uživatelských profilů na základě analýzy multimediálního obsahu, a vizualizačního nástroje implementovaného jako Jupyter notebook.

Klíčová slova

algoritmický audit, digitální legislativa (DSA), doporučovací algoritmus, mobilní agent, personalizace obsahu, sock-puppet

Annotation (or Summary)

This paper presents the design and implementation of a system for automated algorithmic auditing of social media recommendation feeds. The theoretical section describes key issues associated with recommendation algorithms – including disinformation, psychological effects, and content moderation failures – and frames them within the European Digital Services Act (DSA) regulatory framework, critically examining its shortcomings in enforcing platform transparency. The practical section introduces a three-component system: a mobile agent controlling the Instagram Reels application via the UIAutomator2 library, a predictive system simulating synthetic user profile behaviour based on multimedia content analysis, and a visualisation tool implemented as a Jupyter notebook.

Keywords

algorithmic audit, content personalization, Digital Services Act (DSA), mobile agent, recommendation algorithm, sock-puppet

Obsah

1 Úvod

Sociální sítě se za poslední desetiletí proměnily z komunikačních nástrojů v jednu z nejvlivnějších informačních infrastruktur současnosti. Klíčovou roli v této proměně sehrály doporučovací algoritmy – systémy, které rozhodují o tom, jaký obsah uživatel uvidí, v jakém pořadí a v jaké míře. Jejich vliv přesahuje individuální uživatelskou zkušenost a zasahuje do oblasti veřejného diskursu, duševního zdraví, politické komunikace i ochrany nezletilých.

Tato práce se zabývá algoritmickými systémy sociálních sítí z několika vzájemně provázaných perspektiv. Teoretická část nejprve popisuje dokumentované problémy spojené s užíváním sociálních sítí a zasazuje je do kontextu psychologických mechanismů, na nichž jsou platformy postaveny. Následně se věnuje evropskému regulačnímu rámci v podobě nařízení DSA, jeho požadavkům na transparentnost algoritmů a tomu, jak platformy tyto požadavky v praxi plní. Samostatnou pozornost věnuje metodologickým přístupům k auditu algoritmů a kritické analýze první vlny auditních zpráv, jejichž nedostatky poukazují na strukturální limity stávajícího dohledového rámce.

Praktická část navazuje na tato zjištění a představuje třísložkový systém pro automatizované algoritmické audity doporučovacích feedů. První složku tvoří mobilní agent, který na telefonech s operačním systémem Android autonomně ovládá aplikaci Instagram Reels a simuluje přirozené interakční vzorce reálných uživatelů. Druhou složkou je prediktivní systém, který rozhoduje o konkrétní interakci s videem na základě analýzy multimediálního obsahu videa, definovaného uživatelského profilu a obecných vzorců chování uživatelů v prostředí doporučovacích feedů. Třetí složkou je vizualizační nástroj v podobě Jupyter notebooku převádějící auditní data do interpretovatelných grafů.

2 POUŽITÉ TECHNOLOGIE

Jako hlavní programovací jazyk pro vývoj všech tří klíčových složek systému byl zvolen Python, a to především díky jeho rozsáhlému ekosystému knihoven pro automatizaci, zpracování multimédií a analýzu dat. Tato volba umožňuje efektivní integraci pokročilých nástrojů pro multimodální analýzu obsahu a sémantickou interpretaci dat, přičemž ty nejzásadnější z nich jsou podrobněji specifikovány v následující sekci. Pro účely správy celého ekosystému, konfiguraci auditních scénářů a následný export dat do vizualizačního nástroje bylo vytvořeno uživatelské rozhraní postavené na frameworku Next.js.

2.1 Knihovny

2.1.1 UIAutomator2

UIAutomator2 je framework pro automatizované testování uživatelského rozhraní aplikací na platformě Android. Umožňuje simulovat interakce uživatele s aplikací – například klikání, psaní textu, scrollování nebo ovládání systémových prvků. Na rozdíl od některých jiných testovacích nástrojů dokáže UIAutomator2 pracovat i mimo samotnou testovanou aplikaci, což znamená, že může interagovat s celým systémem zařízení (např. notifikace, systémová nastavení apod.). Framework využívá identifikaci UI prvků pomocí atributů, jako jsou text, resource-id nebo class name, a umožňuje tak přesné cílení na konkrétní komponenty rozhraní. Díky tomu je vhodný pro automatizaci end-to-end testů mobilních aplikací a zajištění jejich správné funkčnosti z pohledu uživatele.

2.1.2 Android Debug Bridge

ADB je oficiální komunikační protokol vyvinutý společností Google pro interakci mezi vývojářským prostředím a zařízením s operačním systémem Android. Funguje na principu client-server architektury, kdy ADB démon běžící na zařízení komunikuje s klientem na pracovní stanici prostřednictvím USB nebo TCP/IP. ADB poskytuje rozhraní pro přístup k systémovému shellu zařízení, k souborovému systému, k logovacímu subsystému (logcat) i k automatizovanému ovládání UI vrstvy, na němž je postavena knihovna UIAutomator2. V navrženém systému ADB tvoří základní komunikační kanál mezi pracovní stanicí, na níž je spuštěn agent, a fyzickým mobilním zařízením, které simuluje reálného uživatele.

2.1.3 Pydantic

Pydantic je knihovna pro validaci dat a správu nastavení v jazyce Python, využívající Python type hints. Automaticky validuje a konvertuje vstupní data do definovaných datových struktur, přičemž poskytuje jasné chybové zprávy při neshodě. Pydantic je úzce integrována s FastAPI a tvoří základ pro validaci API požadavků a odpovědí, čímž zajišťuje typovou bezpečnost a automatickou generaci schémat.

2.1.4 yt-dlp

yt-dlp je open-source nástroj pro stahování videí z více než 1000 podporovaných platforem, včetně YouTube, TikTok a Instagram. Představuje aktivně udržovaný fork původního youtube-dl s vylepšenou podporou moderních platforem a rychlejšími aktualizacemi. Knihovna umožňuje extrakci metadat, stahování specifických časových segmentů videa, volbu kvality a formátu a poskytuje Python API pro programovou kontrolu. V prediktivním systému slouží pro stahování video obsahu k následné analýze bez nutnosti manuálního zásahu.

2.1.5 OpenCV

OpenCV (cv2) je open-source knihovna pro počítačové vidění a zpracování obrazu, původně vyvinutá společností Intel. Poskytuje rozsáhlou sadu algoritmů pro detekci objektů, rozpoznávání vzorů, segmentaci obrazu a práci s videem. V prediktivním systému je využívána pro extrakci jednotlivých snímků (frames) z video souborů, což umožňuje vizuální analýzu obsahu bez nutnosti zpracovávat celé video v reálném čase.

2.1.6 ShazamAPI

ShazamAPI je neoficiální Python wrapper pro Shazam API, umožňující rozpoznávání hudby z audio souborů. Knihovna dokáže identifikovat název skladby, interpreta a další metadata na základě audio otisku. V prediktivním systému slouží pro detekci známé hudby ve video obsahu, což představuje jeden z klíčových faktorů ovlivňujících uživatelské rozhodování při konzumaci krátkých videí.

2.1.7 Pandas

Pandas je open-source knihovna určená k analýze a manipulaci s daty. Poskytuje výkonné datové struktury – především DataFrame a Series – které umožňují efektivní práci s tabulkovými i časovými daty. Knihovna podporuje načítání dat z různých formátů (CSV, parquet, JSON, SQL aj.), jejich čištění, transformaci, agregaci a vizualizaci. pandas je navržena s důrazem na výkon a flexibilitu a tvoří základ většiny datově analytických pracovních postupů v ekosystému Pythonu.

2.1.8 NumPy

NumPy je open-source knihovna pro vědecké výpočty. Jejím ústředním prvkem je výkonné n-rozměrné pole (ndarray), který umožňuje efektivní ukládání a operace s homogenními daty. Knihovna poskytuje rozsáhlou sadu matematických funkcí pro lineární algebru, Fourierovu transformaci a další numerické operace. NumPy je optimalizována pro výkon díky internímu využití jazyka C.

2.1.9 Matplotlib

Matplotlib je komplexní open-source knihovna pro tvorbu statických, animovaných i interaktivních vizualizací. Jejím základním rozhraním je modul pyplot, který poskytuje intuitivní API inspirované prostředím MATLAB. Knihovna podporuje širokou škálu typů grafů – od základních spojnicových a sloupcových grafů až po složité víceosé a 3D vizualizace. Matplotlib nabízí vysokou míru přizpůsobení výstupu a podporuje export do mnoha formátů včetně PNG, PDF či SVG. Je úzce integrována s knihovnami NumPy apandas.

2.1.10 Seaborn

Seaborn další z knihoven pro vizualizaci dat, postavená nad knihovnou Matplotlib. Poskytuje vysokoúrovňové rozhraní pro tvorbu esteticky propracovaných grafů. Seaborn je navržena pro těsnou spolupráci s datovými strukturami knihovny pandas a nabízí specializované typy grafů pro explorativní analýzu dat – například tepelné mapy, houslové grafy, párové grafy aj. Součástí knihovny jsou také předpřipravená barevná schémata a témata, která zajišťují konzistentní a publikačně kvalitní výstup.

2.1.11 FastAPI

FastAPI je webový framework pro jazyk Python určený pro tvorbu výkonných API rozhraní. Využívá asynchronní vzor (ASGI) a pokročilé typové anotace jazyka Python ve spojení s knihovnou Pydantic, čímž zajišťuje automatickou validaci vstupů, generování OpenAPI dokumentace a typovou bezpečnost komunikace. V navrženém systému tvoří FastAPI backendovou vrstvu prediktivního systému – konkrétně zpřístupňuje veřejný endpoint pro výpočet doporučené uživatelské interakce a interní endpointy pro autentizaci, správu agentů a správu auditovacích období.

2.1.12 Next.js

Next.js je produkční framework postavený nad knihovnou React, určený pro tvorbu serverově renderovaných i klientských webových aplikací. V navrženém systému slouží Next.js jako frontendová vrstva administračního rozhraní prediktivního systému; zajišťuje uživatelské rozhraní pro vytváření a konfiguraci agentů, správu auditovacích session a export shromážděných auditních dat do vizualizačního nástroje.

3 TEORETICKÁ ČÁST

3.1 Problémy sociálních sítí

V rámci EU využívá internet každý den 96 % mladých lidí ve věku 16–29 let a 84 % z nich se aktivně účastní sociálních sítí. (European Commission 2025) Každá z těchto platforem deklaruje závazné zásady komunitního chování – přesto se uživatelé pravidelně setkávají s obsahem, který by dle zásad neměl být dostupný. Tato sekce popisuje, o jaký obsah se jedná a proč k jeho výskytu dochází.

3.1.1 Dezinformace a manipulativní obsah

Nejrozšířenější kategorií škodlivého obsahu jsou dezinformace. Na týdenní bázi se s dezinformacemi setkává 44,2 % adolescentů – více než s jakoukoli jinou sledovanou kategorií online hrozeb. (Lahti et al. 2024) Jejich prevalence není náhodná: nepravdivé zprávy se šíří na Twitteru šestkrát rychleji než pravdivé a jsou o 70 % více sdíleny. Během amerických voleb 2020 dosahoval falešný obsah na Facebooku šestinásobně vyššího počtu interakcí oproti faktickému obsahu. (Jalli 2024) Tento jev je systémový – platformy jsou konstruovány tak, aby maximalizovaly reakce uživatelů, emocionálně nabitý obsah (pravdivý i nepravdivý) přirozeně generuje více reakcí.

Za zvlášť nebezpečnou podkategorii lze považovat politicky motivované dezinformace. Například šetření organizace India Civil Watch International odhalilo, že Meta schválila politické inzeráty obsahující otevřenou protimuslimskou nenávist, konspirace namířené proti opozičním politikům a výzvy k násilí – a to navzdory veřejně deklarované politice potírání nenávistných projevů. (Jalli 2024)

3.1.2 Obsah škodlivý pro duševní zdraví

Samostatnou kategorií je obsah, jehož škodlivost není primárně spjata s nepravdivostí, ale s psychologickým účinkem na příjemce. Doporučovací algoritmy mohou děti a adolescenty postupně vystavovat nevhodným či škodlivým materiálům; tento mechanismus je o to závadnější, neboť platformy jsou navrženy primárně pro dospělé uživatele. (Joint Research Centre 2025) Výzkum zahrnující přibližně 500 dětí z celého světa ukázal, že mladí uživatelé

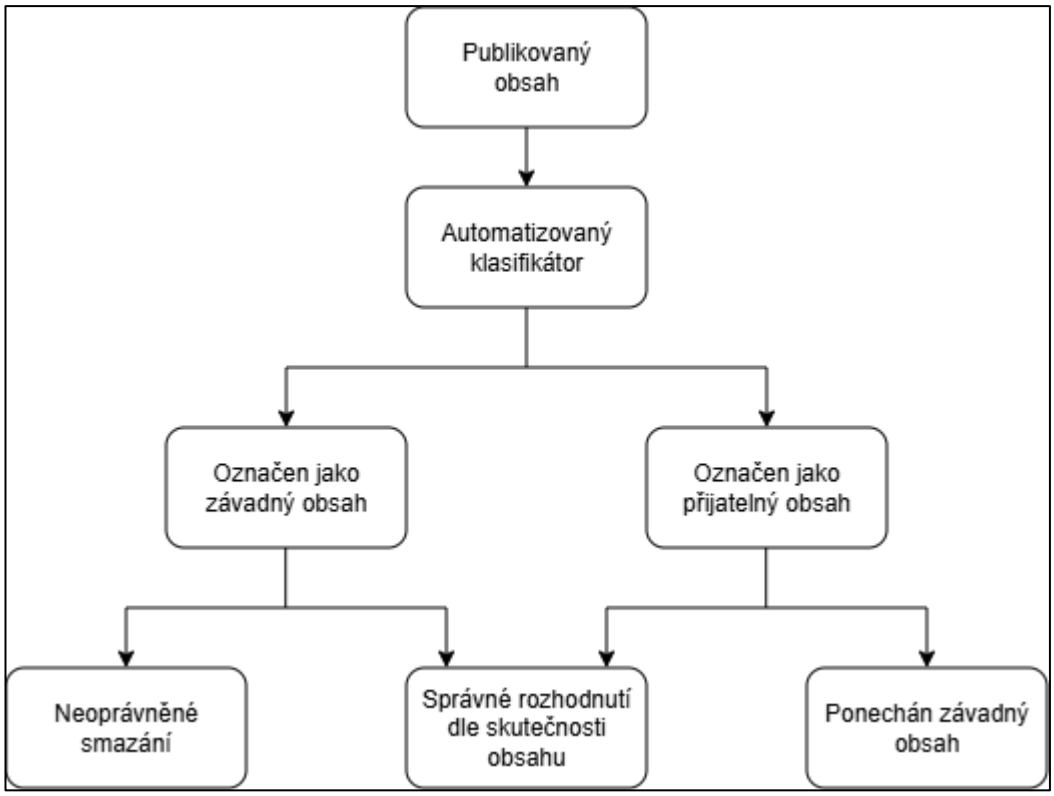
sami popisují platformy jako prostředí, v němž algoritmy vytvářejí tzv. „králičí nory“ depresivního, sebevražedného nebo sebepoškozujícího obsahu. (Goodman 2025)

Obsah vyvolávající tlak na tělesný vzhled je druhou nejčastěji každodenně vídanou hrozbou – setkává se s ním denně 9,1 % adolescentů. (Lahti et al. 2024) Jde primárně o videa propagující nerealistické tělesné ideály, nelegální obsah, diety nebo sebepoškozující návyky v oblasti stravování, který platformy formálně zakazují, avšak jeho moderace je v praxi nekonzistentní.

Metaanalytické studie opakovaně potvrzují asociaci mezi expozicí vůči obsahu zaměřenému na sebepoškozování a zvýšeným rizikem sebevražedného chování u mladých lidí, přičemž tento efekt je výraznější u dívek. (Agyapong-Opoku et al. 2025) Specifickým problémem je, že doporučovací systémy mohou tento typ obsahu algoritmicky zesilovat u právě těch uživatelů, kteří jsou nejvíce zranitelní.

3.1.3 Selhání moderace jako systémový problém

Existence výše popsaného obsahu není důsledkem absence pravidel, ale jejich nedostatečného vymáhání. Platformy se při moderaci ve stále větší míře spoléhají na automatizované systémy strojového učení, které nevyhnutelně generují jak falešně pozitivní výsledky (přijatelný obsah je odstraněn), tak falešně negativní (závadný obsah zůstane dostupný). (Agyapong-Opoku et al. 2025) Zpráva Social Media Safety Index 2024 konstatuje přetrvávající selhání hlavních platforem při potírání nenávistných projevů a dezinformací navzdory deklarovaným zásadám – přičemž zároveň dochází k neoprávněnému mazání legitimního obsahu marginalizovaných skupin, jak ukazuje obrázek 1. (Kozyreva et al. 2022)



Obrázek 1: Diagram rozhodování o obsahu

Toto strukturální napětí mezi rozsahem moderace a její přesností je jedním z ústředních problémů, jimž se věnují následující kapitoly – zejména v kontextu transparentnosti doporučovacích algoritmů (3.5.1) a metodologických přístupů k jejich auditu (3.6.2).

3.2 Teoretický popis fungování algoritmů

Doporučovací algoritmy jsou dnes základní infrastrukturou každé velké sociální sítě – určují, který obsah se uživateli zobrazí, v jakém pořadí a v jaké míře. Na technické úrovni pracují tyto systémy zpravidla se třemi základními přístupy nebo jejich kombinací. Kolaborativní filtrování (collaborative filtering) vychází z podobnosti chování uživatelů navzájem: pokud dva uživatelé v minulosti interagovali s podobným obsahem, systém jednomu z nich doporučí obsah, který zaujal toho druhého. Filtrování na základě obsahu (content-based filtering) naopak analyzuje vlastnosti samotného obsahu – tagy, délku videa, téma, formát – a porovnává je s uživatelským profilem sestaveným z jeho dosavadní aktivity. Moderní platformy jako TikTok nebo YouTube kombinují oba přístupy do hybridních modelů, které doplňují i kontextové signály: typ zařízení, denní dobu, geolokaci nebo aktuálně trendující obsah. (Zhou 2024) Klíčovou optimalizační proměnnou těchto systémů je engagement – čas strávený sledováním, počet interakcí, míra

dokoukání videa. Algoritmus se učí predikovat, jaký obsah maximalizuje tuto metriku, a tomu přizpůsobuje doporučení v reálném čase.

Výsledkem je inherentní napětí mezi **personalizací a diverzitou obsahu**. Čím přesněji algoritmus optimalizuje na prokázané preference uživatele, tím více se obsah jeho feedu zužuje kolem opakujících se témat a formátů. Rozsáhlá obsahová analýza platforem ukázala, že hyper-personalizace snižuje diverzitu zobrazovaného obsahu o přibližně třetinu, přičemž intenzivní uživatelé vykazují výraznou tematickou stagnaci – tedy situaci, kdy feed přestává přinášet nové informační podněty. (Starke et al. 2025) Tento efekt je přímým důsledkem designu, nikoliv vedlejším produktem: systémy explicitně optimalizující na engagement zpravidla diverzitu oslabují, zatímco systémy, které zahrnují metriky rozmanitosti nebo prvky náhody, tento efekt zmírňují.

S tím úzce souvisejí pojmy **filter bubble** a **echo chamber**, které se v odborné i veřejné diskusi používají k popisu situace, kdy uživatel je algoritmicky izolován od obsahu neodpovídajícího jeho stávajícím preferencím. Ačkoliv oba termíny popisují podobný jev informační uzavřenosti, liší se původem: filter bubble označuje uzavřenost způsobenou algoritmickým tříděním bez vědomé volby uživatele, zatímco echo chamber odkazuje spíše na prostředí aktivně formované uživatelem samotným – sledovanými účty, sociálními sítěmi a záměrnou volbou zdrojů. (Ross Arguedas et al. 2022)

Empirická literatura na obě hypotézy pohlíží opatrně: přehledová studie Reuters Institute uvádí, že naprostá většina empirických výzkumů nenalezla algoritmicky způsobenou informační uzavřenost v takové míře, jakou naznačuje veřejná debata, a že uživatelé v průměru konzumují informačně rozmanitější obsah prostřednictvím sociálních sítí než skrze tradičnější offline kanály. (Ross Arguedas et al. 2022)

Systematická recenze zaměřená na mladé uživatele z roku 2025 naopak potvrzuje, že algoritmické systémy strukturálně zesilují ideologickou homogenitu a omezují diverzitu pohledů – a upozorňuje na to, že efekt závisí výrazně na designu konkrétní platformy a na míře digitální gramotnosti uživatele. (Starke et al. 2025) Spor o rozsah těchto jevů zůstává v akademické obci otevřený; co se však ukazuje jako nesporné, je to, že doporučovací systémy jsou dostatečně výkonné na to, aby informační prostředí jednotlivce formovaly – a že transparentnost jejich nastavení je pro posouzení tohoto vlivu nezbytná.

3.3 Psychologické aspekty užívání a rizika algoritmů

Aby bylo možné porozumět tomu, proč je transparentnost algoritmů tak zásadní, je třeba pochopit mechanismy, které uživatele na platformách udržují – a co se stane, když jsou tyto mechanismy zneužity.

Výzkum zaměřený na TikTok prokázal, že uživatel, který začne interagovat s určitým typem extremistického obsahu, může být algoritmem nasměrován do výrazně radikálnějšího informačního prostředí v průběhu pouhých dvou hodin konzumace – a to i v případě, že daný obsah formálně porušuje pravidla platformy. (Shin a Jitkajornwanich 2024) Studie z roku 2024 zaměřená na volby v Rumunsku ukázala, že doporučovací algoritmus TikToku, podporovaný sítěmi neautentických účtů, výrazně posílil viditelnost marginálního extremistického kandidáta. (Solarova et al. 2026) Akademici v oblasti politické vědy upozorňují, že algoritmy přispívají k polarizaci skrze mechanismus zesilování emocionálně nabité a dělící komunikace, protože ta generuje vyšší engagement než obsah neutrální. (Wang a Wang 2025)

Netransparentnost přitom vytváří strukturální mocenskou asymetrii: platforma disponuje kompletním obrazem o dopadu svých systémů, zatímco uživatelé, regulátoři i výzkumníci jsou odkázáni na nepřímé metody pozorování. Tato asymetrie je zvláště riziková v kontextu voleb, informačních operací a cílení zranitelných skupin. Jak poznamenal výzkum publikovaný v časopise *Perspectives on Psychological Science*, algoritmické mechanismy a sociální hybatele vytvářejí zpětnovazební smyčku, v níž je jejich vzájemný vliv prakticky neoddělitelný – a tedy obtížně měřitelný bez přístupu k interním datům platformy. (De et al. 2025)

3.4 Prevence a legislativy (DSA)

Reakcí Evropské unie na problémy popsané v předchozí kapitole je zejména nařízení o digitálních službách – **Digital Services Act (DSA)**, které představuje dosud nejambicióznější pokus o systémovou regulaci online platforem. DSA si klade za cíl ukončit éru, v níž si technologické společnosti regulovaly samy sebe – stanovovaly vlastní pravidla pro moderaci obsahu a vydávaly transparentní zprávy o svém boji s dezinformacemi, které bylo prakticky nemožné pro třetí strany ověřit. (European Parliament 2024)

Vznik, struktura a právní rámec DSA

DSA byl podepsán do zákona dne 19. října 2022 a vstoupil v platnost 16. listopadu 2022; nařízení bylo přijato Evropským parlamentem 539 hlasy pro, 54 proti a se 30 zdržením se hlasování. (Parliament 2024) Dne 25. srpna 2023 začal DSA platit pro tzv. Very Large Online Platforms (VLOPs) a Very Large Online Search Engines (VLOSEs) a od 17. února 2024 se jeho ustanovení rozšířila na všechny ostatní poskytovatele online služeb v EU. (EU 2025)

DSA uplatňuje proporcionální přístup – povinnosti jsou úměrné velikosti a dopadu platformy. Největší online platformy, tedy ty s více než 45 miliony měsíčních uživatelů v EU, nesou nejvyšší odpovědnost a nejrozsáhlejší povinnosti, zatímco mikropodnikům a malým firmám jsou požadavky zmírněny. (European Parliament 2026) Jde o rozsáhlé nařízení obsahující 93 článků a 156 recitálů s širokým dosahem – platí i pro platformy se sídlem mimo EU, pokud nabízejí služby uživatelům na území EU. (Navea 2024)

3.4.1 Vymáhání a sankce

Vymáhání DSA je sdíleno mezi Evropskou komisí a národními orgány. Komise je výhradně příslušná pro dohled nad největšími platformami, zatímco národní orgány – tzv. Digital Services Coordinators (DSC) – dohlížejí na menší poskytovatele ve svém členském státě. (European Commission 2025). Za nedodržení povinností mohou platformám hrozit pokuty až do výše 6 % jejich celosvětového ročního obratu. (Vcard 2024) Tento mechanismus byl v praxi již aktivován: platformě X byla uložena pokuta 45 milionů EUR za nesoulad jejího reklamního repozitáře s požadavky DSA. (European Parliament 2026)

DSA zavádí celou sadu konkrétních nástrojů zaměřených na ochranu uživatelů, z nichž nejvýznamnější se týkají transparentnosti, moderace obsahu a ochrany zranitelných skupin.

3.4.2 Transparentnost doporučovacích systémů a reklamy

DSA ukládá provozovatelům online platforem povinnost zajistit větší transparentnost a kontrolu nad obsahem, který se zobrazuje v uživatelských feedech. Uživatelé tak mohou zjistit, na jakém základě platformy obsah seřazují, a mají právo odmítnout personalizovaná doporučení – VLOPs musí nabídnout možnost vypnout personalizovaný obsah. Tato povinnost přímo reaguje na netransparentnost algoritmického doporučování.

3.4.3 Právo na odvolání a mimosoudní řešení sporů

Platformy jsou nově povinny uživatelům zdůvodnit každé rozhodnutí o moderaci a umožnit jeho napadení. Od roku 2024 uživatelé v EU podali prostřednictvím interních mechanismů platforem více než 165 milionů odvolání proti moderačním rozhodnutím VLOPs a VLOSEs, přičemž téměř 30 % z nich vedlo ke změně rozhodnutí. V první polovině roku 2025 mimosoudní orgány přezkoumaly přes 1 800 sporů týkajících se obsahu na Facebooku, Instagramu a TikToku a v 52 % uzavřených případů rozhodnutí platformy zrušily. (European Parliament 2026)

3.4.4 Ochrana nezletilých jako priorita DSA

Zvláštní pozornost věnuje DSA ochraně dětí a mladistvých, kteří jsou, jak bylo popsáno v kapitole 3.1, vůči škodlivému obsahu nejzranitelnější skupinou. Pokyny Komise k ochraně nezletilých pod článkem 28 DSA se vztahují na jakoukoliv službu, která je nezletilými využívána nebo u níž lze jejich využívání předpokládat – nestačí tedy, aby platforma formálně deklarovala, že je určena pouze dospělým, pokud k ní mají mladí uživatelé fakticky přístup. (CADE Project 2025)

3.4.5 Ověřování věku

Jedním z nejdiskutovanějších nástrojů je povinnost ověřování věku. Komise připravuje technické řešení respektující soukromí uživatelů, které by umožnilo potvrdit věk bez sdílení dalších osobních údajů; toto řešení je v současné době v pilotní fázi a je testováno ve spolupráci s členskými státy, platformami a koncovými uživateli. (European Parliament 2026) Komise přitom navrhuje vrstvený přístup: platformy s vysokým rizikem by měly implementovat plnohodnotné ověřování věku, platformy se středním rizikem odhad věku a platformy s nízkým rizikem nemusí přijímat žádná věková opatření. (Allen 2025; Parliament 2024)

3.4.6 Safety-by-design a doporučovací systémy pro nezletilé

Platformy jsou povinny zajistit, aby doporučovací systémy nevystavovaly nezletilé škodlivému nebo nelegálnímu obsahu. Nezletilí musí mít možnost resetovat svůj feed, upravit obsahové preference a porozumět tomu, proč je jim konkrétní obsah doporučován; jako výchozí nastavení musí být dostupná možnost doporučování nezaložená na profilování. (Hogan Lovells 2025)

Evropská komise v roce 2024 zahájila šetření proti Facebooku a Instagramu pro podezření z nedodržování pravidel DSA v oblasti ochrany nezletilých – konkrétně zkoumá, zda funkce a algoritmy těchto platforem nepodporují návykové chování u dětí a nevtahují je do tzv. rabbit-hole efektu. (European Commission 2025)

3.5 Jak sociální sítě plní/neplní legislativní požadavky

Přestože DSA stanovuje rozsáhlý rámec povinností, jeho praktická implementace odhaluje zásadní mezery mezi deklarovanou a skutečnou transparentností platforem. Klíčovým problémem není absence pravidel, ale to, že platformy volí jejich minimalistický výklad a zůstávají vůči regulátorům i uživatelům do značné míry neprůhledné.

3.5.1 Transparentnost doporučovacích algoritmů v praxi

DSA v článcích 27 a 38 ukládá platformám povinnost vysvětlovat uživatelům, na jakém základě jim je obsah doporučován, a nabídnout alternativu bez personalizace. Analýza prvních auditorských zpráv z roku 2024 nicméně ukazuje, že platformy zvolily úzký výklad těchto ustanovení. Aplikace článku 27 se dosud omezila pouze na plnění požadavku na vysvětlení pro uživatele, zatímco platformy v naprosté většině případů neimplementovaly nástroje skutečné uživatelské kontroly – tento přístup lze označit jako „minimalistický", neboť odpovídá spíše tomu, co je psáno, než kontextu a pojetí nařízení. (DSA Observatory 2024)

Zároveň se ukazuje, že vysvětlení poskytovaná samotnými platformami jsou nespolehlivá. TikTok například přiřadil odůvodnění „okomentoval jsi podobná videa" ke 34 % videí zobrazených účtům, které na platformě dosud žádný komentář nezanechaly. (Solarova et al. 2026)

Problém transparentnosti se projevuje i na úrovni povinných zpráv. Výzkum DSA Transparency Database odhalil, že téměř 90 % moderačních odůvodnění spadá do vágní kategorie „rozsah platformové služby", aniž by poskytovalo jakékoli konkrétní informace o skutečném důvodu zásahu. (Roy 2025) Jednotlivé platformy si navíc vytvořily vlastní kategorizaci obsahu, takže zprávy jsou vzájemně obtížně srovnatelné. (Ohnesorge 2025) Analytici shrnují situaci tak, že platformám se daří prostřednictvím databáze demonstrovat formální soulad, i když v praxi může být jejich transparentnost nedostatečná. (Roy 2025)

3.5.2 Problémy, které sociální sítě aktivně neřeší

Nad rámec nedostatečné transparentnosti existují oblasti, které platformy neřeší ani při vědomém porušování vlastních zásad. Evropská komise v roce 2024 zahájila formální řízení proti Facebooku, Instagramu a TikToku mj. pro porušení pravidel transparentnosti algoritmů a ochrany nezletilých. (Iyer 2025) V prosinci 2025 byl platformě X uložena pokuta 120 milionů EUR za porušení transparentních povinností v oblasti reklamy a designu rozhraní. (MediaLaws 2025)

Klíčovým strukturálním problémem, který DSA sám o sobě neřeší, je skutečnost, že doporučovací algoritmy jsou pro vnější pozorovatele prakticky neprozkoumatelnými systémy. Interní dokumenty zveřejněné whistleblowerkou Frances Haugenovou v roce 2021 odhalily, že doporučovací systém služby Meta záměrně zesiloval polarizující a škodlivý obsah proto, aby maximalizoval engagement uživatelů. Podobné obavy byly vzneseny v souvislosti s algoritmem sociální sítě TikTok a její For You Page, u níž bylo prokázáno, že zranitelným uživatelům začal doporučovat obsah spojený s poruchami příjmu potravy během několika minut od vytvoření účtu. (Solarova et al. 2026)

3.6 Auditování algoritmů: metodologie, stav a nedostatky

3.6.1 Algoritmické auditování jako možné řešení

Tradiční auditní metodiky vyvinuté pro oblast finančního výkaznictví nejsou pro hodnocení algoritmického chování dostatečné – neumožňují zachytit dynamičnost, kontext a sociální dopady doporučovacích systémů.

Výzkumníci proto upozorňují na potřebu specializovaných přístupů k auditu AI systémů, které kombinují technické, právní i společenskovědní metody. (Ojewale et al. 2025) Algoritmické auditování – tedy systematické testování, hodnocení a dokumentaci chování algoritmických systémů – představuje oblast, která by mohla zaplnit tuto mezeru: umožnila by regulátorům, výzkumníkům i občanské společnosti nezávisle ověřovat, zda doporučovací systémy platforem skutečně odpovídají jejich deklarovaným zásadám i požadavkům DSA.

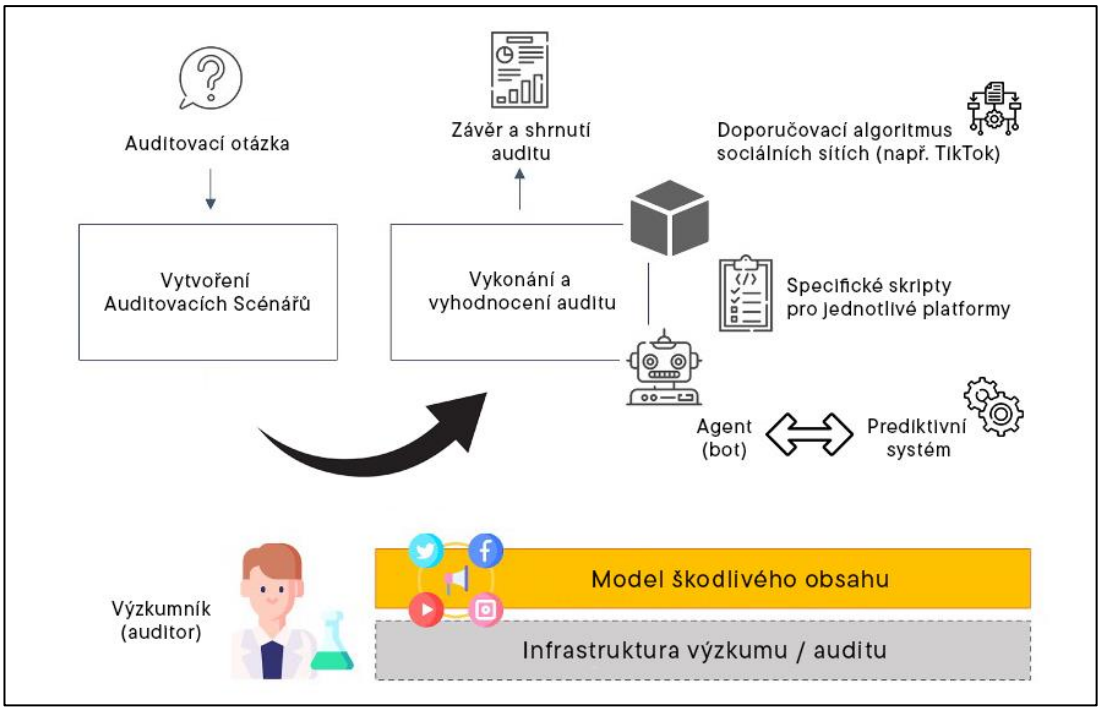
3.6.2 Metodologické přístupy k auditu algoritmů

Aby bylo možné posoudit, zda algoritmy sociálních sítí fungují v souladu s deklarovanými pravidly i právními požadavky, vyvinula vědecká komunita a regulační orgány několik odlišných metodologických přístupů. Každý z nich nabízí jiný pohled na chování systémů, které jsou z podstaty neprůhledné – a každý s sebou nese vlastní omezení.

Nejstarším a nejjednodušším nástrojem je **scraping**, tedy automatizované stahování veřejně dostupných dat přímo z rozhraní platformy. Umožňuje popisnou analýzu obsahu a základní korelační studie, je však citlivý na změny v designu platformy a obtížně zachycuje personalizované chování algoritmů, neboť výsledky se pro každého uživatele liší. (Shin a Jitkajornwanich 2024)

Sofistikovanějším nástrojem je **API audit**, při němž výzkumníci přistupují k datům přes programovací rozhraní, která platforma zpřístupňuje. Oproti scrapingu umožňuje strukturovanější a opakovaně proveditelný sběr dat vhodný ke komparativním studiím. Zásadní nevýhodou však je, že platforma sama kontroluje, jaká data přes API poskytuje, a může tudíž selektivně omezit přístup k těm nejcitlivějším – nebo API po zahájení výzkumu uzavřít. Dobrým příkladem je Twitter/X, který v roce 2023 drasticky omezil přístup k výzkumné API, čímž zablokoval desítky dlouhodobých výzkumných projektů. (Ada Lovelace Institute 2021)

Nejrozšířenějším metodologickým nástrojem pro studium algoritmické personalizace je tzv. **sock-puppet audit**. Výzkumníci vytvářejí sady automatizovaných falešných účtů – „loutkových uživatelů“ – s přesně definovanými profily a sledují, jak algoritmus na jejich chování reaguje. Metodologii sock-puppet auditu znázorňuje obrázek 2. (Srba et al. 2025) Klíčovou výhodou je kontrolovatelnost: proměnné jako věk účtu, sledované zdroje, politická orientace nebo typ konzumovaného obsahu lze přesně nastavit, zatímco u skutečných uživatelů by tyto faktory nebyly možné oddělit. Výzkum z konference FAccT 2025 nasadil 120 sock-puppet účtů na platformě X rozdělených do čtyř skupin podle politické orientace a zdokumentoval, že algoritmus platformy systematicky zesiluje obsah odpovídající politickému přesvědčení uživatele a nové účty (bez sledovaných profilů) vykazují jako výchozí stav pravicové zkreslení doporučení. (Ye et al. 2025)



Obrázek 2: Metodologie sock-puppet auditu

Sock-puppet audity mají však zásadní metodologické omezení: automatizovaní agenti nemohou plně napodobit lidské chování, protože neklikají spontánně, nepřesouvají kurzor nepravidelně ani nereagují na sociální kontext. Výsledky citlivě závisí také na zdánlivě technických rozhodnutích – počtu sledovaných účtů, délce relace nebo způsobu segmentace uživatelů – přičemž různé volby těchto parametrů mohou vést k odlišným závěrům. (Bouchaud a Ramaciotti 2024) Méně diskutovanou otázkou je také právní rámec: sock-puppet audity zpravidla porušují podmínky použití platformy, i když v Evropě i USA bývají pro výzkumné – nekomerční – účely zpravidla tolerovány. (Ada Lovelace Institute 2021)

Nejnověji se prosazuje koncept **model-based auditing**, který namísto přímé interakce s platformou vytváří její zjednodušenou simulaci, na níž lze testovat algoritmické hypotézy. Tento přístup je nezávislý na platformě a snáze reprodukovatelný, zatím však zůstává ve fázi výzkumné propozice. (Srba et al. 2025)

3.6.3 Dosavadní audity: co bylo prověřeno a co zůstává stranou

První povinná vlna auditů dle článku 37 DSA proběhla ve čtvrtém čtvrtletí roku 2024. Zprávy pokrývající YouTube, Facebook, Instagram a TikTok zpracovaly převážně firmy Velké čtyřky – EY, KPMG, Deloitte a Holistic AI. (Ada Lovelace Institute 2021) Audit se zaměřil primárně na procedurální soulad: zda platforma disponuje interními procesy pro hodnocení systémových rizik, zda má dokumentovány mechanismy moderace a zda splňuje formální požadavky transparentnosti vůči uživatelům. Ve všech těchto oblastech platformy dosáhly průměrné shody přesahující 85 %, přičemž každá z nich nicméně obdržela celkové negativní hodnocení kvůli nesplnění části povinností. (Tech Policy Press 2024)

Výraznějšího pokrytí se dostalo transparentnosti reklamních archivů, kde DSA stanovuje jasné a technicky ověřitelné požadavky, a základním mechanismům pro podávání stížností ze strany uživatelů. Naopak hlubšímu prověření se systematicky vyhýbala oblast **fungování doporučovacích algoritmů jako takových** – tedy otázka, zda algoritmy skutečně fungují tak, jak je platforma popisuje, a zda jsou jejich výstupy v souladu s deklarovanými zásadami ochrany uživatelů. (Solarova et al. 2026)

Zvláštní pozornost si zaslouží analýza DSA Transparency Database, která sleduje hlášení o moderačních rozhodnutích od zahájení provozu v září 2023. Výzkumníci, kteří analyzovali prvních sto dní databáze, odhalili závažné vnitřní rozpornosti: u platformy X (Twitter) se v interní transparentní zprávě platforma hlásí k rozsáhlému využívání automatizovaných systémů strojového učení, přičemž v samotné databázi za sledované období nevykázala ani jedno částečně automatizované rozhodnutí. U TikToku dosahoval rozdíl mezi podílem automaticky odstraněného obsahu v transparentní zprávě a v databázi přes 50procentních bodů. (Trujillo et al. 2025) Tento rozdíl není náhodnou chybou – je strukturálním rysem systému, jehož výstupy vycházejí z dat dodávaných samotnou auditovanou stranou.

3.6.4 Nedostatky: nejednotnost, definice a problém metodologické standardizace

Výzkum publikovaný v roce 2026 zaměřený specificky na první vlnu DSA auditů identifikuje několik vzájemně propojených selhání, která z auditů v jejich nynější podobě činí nedostatečný nástroj pro posouzení skutečného algoritmického chování. (Solarova et al. 2026)

Prvním problémem je **absence závazných definic klíčových pojmů**. DSA ukládá platformám, aby uživatelům vysvětlily „nejdůležitější" parametry doporučovacích systémů „jasným a srozumitelným jazykem" – ale ani jeden z těchto termínů není v regulaci precizně vymezen. Platformy i auditoři si proto vytvářejí vlastní interpretace a měřítka. (Solarova et al. 2026) Výsledkem je, že zprávy jsou navzájem nesrovnatelné: každá platforma si zvolila jiný rámec, jiné kategorie a jiné metriky, takže neexistuje základ pro meziplatformní srovnání. Analýza transparentních zpráv sedmi VLOPs za rok 2024 ukázala, že každá platforma si vytvořila vlastní sadu kategorií pro klasifikaci protiprávního obsahu – formálně odvozených od evropského práva, ale fakticky nesourodých a vzájemně neslučitelných. (Ohnesorge 2025)

Druhým selháním je **metodologická mělkost**. Auditoři přistupují k algoritmickým systémům s nástroji, které byly navrženy pro finanční výkaznictví a IT kontroly statických systémů. Doporučovací algoritmy jsou však dynamické, kontextově závislé a neustále se měnící – a to způsobem, který tradiční auditní přístupy nejsou schopny zachytit. EY ve svých zprávách výslovně uvedla, že nevyjadřuje žádný názor, závěr ani jistotu ohledně návrhu, provozu a monitoringu algoritmických systémů – čímž de facto rezignovala na věcné prověření algoritmů jako takových. Podobné zmínky se objevují i v auditech KPMG a Deloitte. DSA Observatory toto označilo jako „bílou vlajku" – auditor se formálně dovolává technické nemožnosti jako důvodu k vyhnutí se auditní odpovědnosti. (DSA Observatory 2024)

Třetím problémem je **časová nestálost** auditovaných systémů. Výzkum zaměřený na funkce TikToku zdokumentoval, že rozhraní dostupná v době výzkumu zmizela ještě před dokončením studie. (Solarova et al. 2026) Tradiční auditní logika předpokládá, že shoda potvrzená v daném okamžiku zůstává platná do příštího auditu – u algoritmů trénovaných kontinuálně na nových datech to neplatí: systém může být v souladu v době auditu a týden poté nikoli, aniž by kdokoli cokoli vědomě změnil.

Čtvrtým a možná nejzásadnějším nedostatkem je **informační asymetrie**. Auditoři z poradenských firem sice mají přístup k interním datům platforem – na rozdíl od nezávislých

výzkumníků –, ale tento přístup probíhá v podmínkách nastavených samotnou platformou, bez práva na zveřejnění citlivých zjištění a bez možnosti srovnání s jinými platformami. (Srba et al. 2025) Výzkumníci operující ve veřejném akademickém prostoru jsou naopak odkázáni na nepřímé metody (sock-puppeting, scraping, API analýzu) a výsledky, jichž dosahují, jsou platforma schopna zpochybnit odkazem na to, že studie nevychází z interních dat. Článek 40 DSA, který by mohl tuto asymetrii řešit zpřístupněním dat pro akreditované výzkumníky, zůstal k roku 2025 z velké části nevyužit – implementace jeho prováděcích předpisů se protahuje a přístup k datům je zatím výjimkou, nikoliv pravidlem. (Krafft et al. 2024)

Výsledkem je stav, kdy regulace algoritmů existuje, auditní povinnost existuje, ale schopnost auditů zachytit skutečné chování algoritmů zůstává výrazně omezená. Bez standardizovaných definic, bez metodologické shody a bez strukturovaného přístupu výzkumníků k datům platforem nelze spolehlivě určit ani to, co by „správně fungující" algoritmus měl vlastně dělat – a tudíž ani to, kdy selhal.

Toto zjištění přirozeně vyvolává otázku, co by skutečně efektivní regulace algoritmů vyžadovala. Odpověď není jednoznačná: příliš striktní nebo technicky nepřiměřené požadavky mohou přinést vlastní rizika – od zásahů do svobody projevu po vytvoření neúnosné administrativní zátěže pro menší platformy.

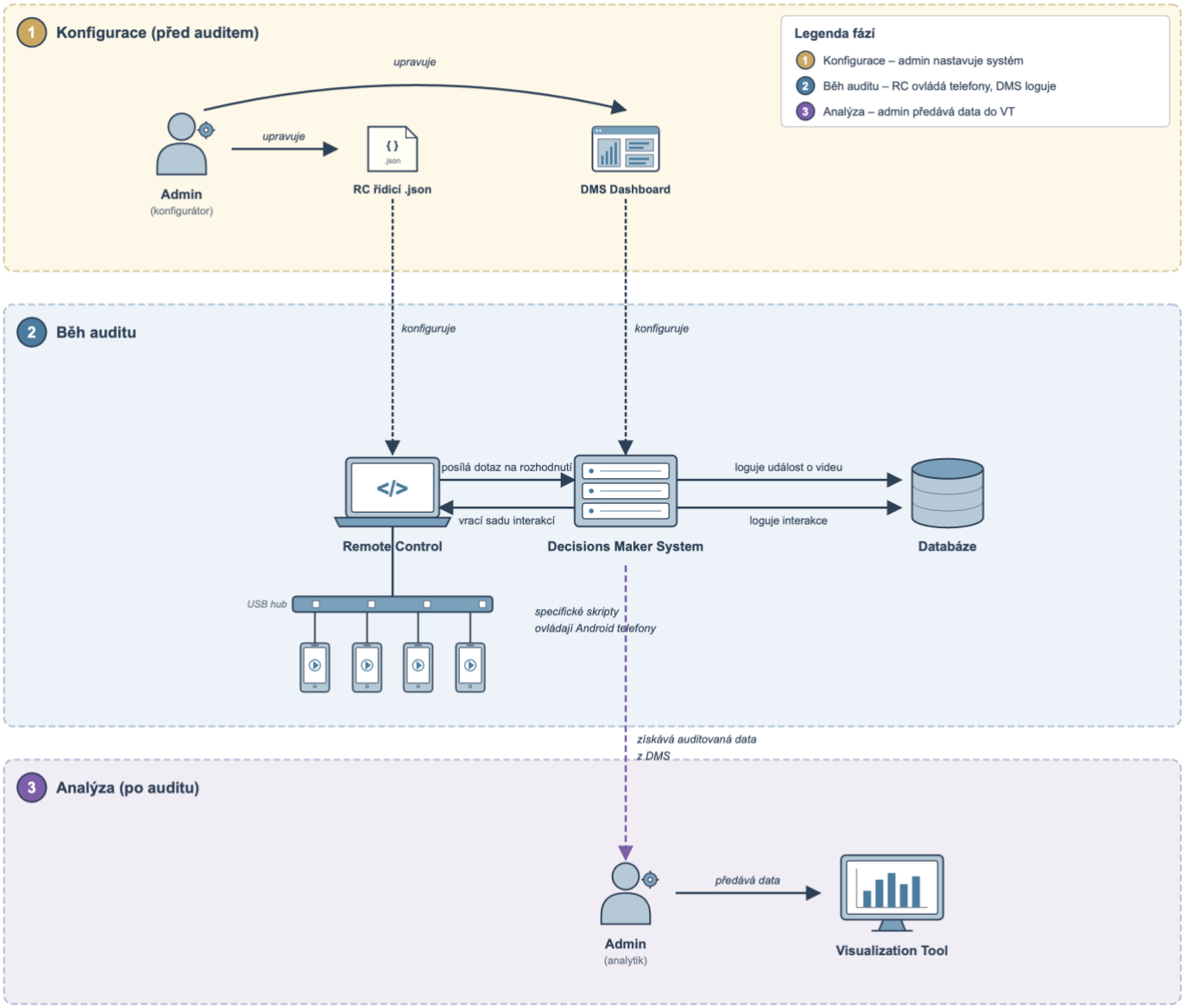
4 CÍLE PRÁCE

Hlavním cílem této práce je navrhnout a realizovat vzorový systém pro automatizované algoritmické audity doporučovacích feedů sociálních sítí (konkrétní provedení pro platformu Instagram Reels), který překonává metodologické nedostatky identifikované v dosavadní auditní praxi – zejména absenci standardizace a nesrovnatelnost výsledků mezi platformami. Aby bylo tohoto cíle dosaženo, je práce rozdělena do tří vzájemně provázaných částí: návrh mobilního agenta schopného autonomně interagovat s prostředím aplikace Instagram (Remote Control = RC), návrh systému simulujícího rozhodování syntetických uživatelských profilů (Decisions Maker System = DMS) a návrh systému pro vizualizaci a interpretaci výsledných auditních dat (Visualization Tool (VT). Schéma kompletní myšlenky auditu je znázorněno obrázkem 3.

Díky architektuře systému je uzpůsobení pro jinou sociální síť již relativně jednoduché, je třeba pouze pozměnit logiku hierarchie Remote Control systému (při zachování vlastností vzoru). Ostatní moduly systému mohou zůstat zachovány. Stejně tak pokud se v průběhu výzkumu doporučovacích algoritmů (pomocí tohoto nástroje) ustálí metriky pro hodnocení experimentů a způsoby jejich zobrazování, je možné je využívat i pro další sociální sítě a postupně zavést standardizaci těchto metrik.

5 Realizace

V rámci realizace byly úkoly všech autorů rozděleny na již zmíněné subsystémy. Zde uvádíme kompletní schéma systému, o jednotlivých částech informujeme dále.



Obrázek 3: Schéma celého systému pro auditování

5.1 Remote control (RC)

Hlavním úkolem subsystému Remote Control je **autonomně ovládat fyzická zařízení s OS Android** a věrohodně na nich simulovat chování reálného uživatele aplikací Instagram Reels a TikTok. Komerční sociální platformy nenabízejí veřejné API pro auditní účely, takže přístup k feedu i k metadatům jednotlivých videí je možný pouze přes vrstvu uživatelského rozhraní mobilní aplikace, doplněnou o privilegovaný přístup k privátnímu úložišti a paměti aplikace, který vyžaduje rootnuté zařízení.

Sekundárním požadavkem je **odolnost proti mechanismům platforem pro detekci syntetických účtů:** simulované interakce musejí být časově nepravidelné, ovládací gesta variabilní a v rámci jedné fyzické instance zařízení nesmějí korelovat tak, aby umožnily detekci automatizace. Nesplnění tohoto požadavku nevede pouze k blokaci účtu, ale potenciálně k vystavení účtu pozměněnému doporučovacímu modelu, což by zcela zdiskreditovalo výsledky auditu.

5.2 Realizace RC

Remote control je realizován jako sada Python skriptů spuštěných na pracovní stanici (notebooku), která prostřednictvím protokolu **Android Debug Bridge (ADB)** a knihovny **UIautomator2** ovládá několik současně připojených android zařízení. Ta jsou fyzicky připojena přes aktivní USB hub, který umožňuje paralelní provoz více zařízení. Každé zařízení je v rámci RC adresováno svým sériovým číslem, které admin před auditem ověří a zapíše ho do JSON souboru. Podle tohoto unikátního čísla se definuje chování zařízení v rámci auditního scénáře. Jedno zařízení tak může mít v oblibě např. politická videa, druhé potom obsah se zvířaty.

Pro získání privilegovaného přístupu k vnitřnímu stavu aplikací byla zařízení rootována nástrojem Magisk. Ten využívá tzv. systemless přístup – rootovací vrstva je aplikována pouze do boot oblasti zařízení, aniž by byla narušena hlavní systémová oblast. Takový přístup zajišťuje, že rootnuté zařízení projde i u aplikací s vyšší citlivostí na integritu systému, mezi které patří například Instagram. Po fyzickém propojení a inicializaci ADB spojení skript přečte řídicí JSON, navrženým způsobem si od **Decisions Maker Systemu** (DMS) vyžádá unikátní video ID, kterým budou autentizovány všechny následné požadavky, a spustí na cílovém

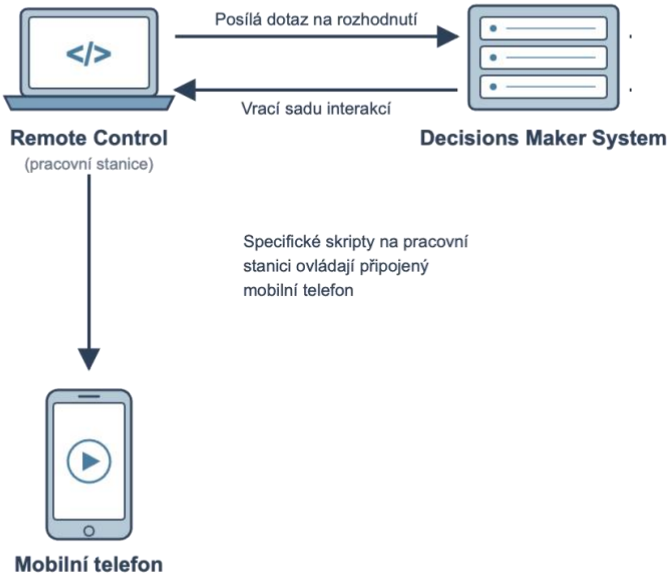
zařízení odpovídající aplikaci (Instagram Reels nebo TikTok). Po jejím nastartování otevře aplikaci v prostředí krátkých videí, simulační smyčka začne od prvního zobrazeného videa.

Klíčovou částí smyčky je **extrakce metadat zobrazeného videa**. Část metadat (popisek, hashtagy, viditelné engagement metriky, autor) se získává z UI vrstvy přes UIAutomator2, který nad pohledovou hierarchií aplikace umožňuje dotazy podle resource-id, class, či textového obsahu prvků. Stabilní identifikátor videa (video ID) v UI dostupný není, a protože slouží jako primární klíč pro zápisu do logů a navazující analýzu, je to komplikace. Mechanismus jeho extrakce prošel během vývoje zásadní revizí, která ilustruje povahu inženýrských kompromisů při auditu uzavřených platforem.

Původní implementace video ID získávala dotazem do interní SQLite databáze aplikace, v níž si Instagram lokálně cachuje metadata zobrazovaných příspěvků. Tato strategie se ukázala jako nedostatečně spolehlivá: aplikace nezapisuje záznamy do SQLite v okamžiku zobrazení videa, ale dávkově, v clusterech v rámci synchronizačních cyklů. ID právě zobrazeného videa tak v databázi typicky chybí přesně v ten okamžik, kdy ho RC potřebuje, a velká část datových bodů by se ve výsledné stopě auditu ztratila.

Aktuální implementace proto extrakci přesouvá na nižší úroveň – video ID není čteno z perzistentního úložiště, nýbrž přímo z paměti RAM běžícího procesu aplikace prostřednictvím rozhraní /proc/<pid>/mem dostupného pouze v rootovaném prostředí. V okamžiku zobrazení videa je ID v paměti vždy přítomno, čímž bylo dosaženo téměř stoprocentní zachycenosti datových bodů. Cenou za toto zlepšení je vyšší vazba implementace na konkrétní verzi cílové aplikace – paměťové rozložení relevantních objektů se mění při každé větší aktualizaci a vyžaduje revalidaci.

Extrahovaná metadata RC serializuje a odešle jako HTTP požadavek na endpoint DMS. Odpověď DMS specifikuje konkrétní akci nebo sekvenci akcí, kterou má RC vykonat (přehrát celé video, lajknout, sdílet, přejít dál, otevřít komentáře, doplnit komentář). RC tuto specifikaci převádí na **fyzická UI gesta** s nepravidelnými časovými rozestupy a se zakřivenými trajektoriemi swipů – cesta prstu po obrazovce není přímka, ale křivka s drobnými šumovými odchylkami, což odpovídá běžnému lidskému ovládání dotykového zařízení. Mezi gesty jsou navíc vkládány náhodné prodlevy odpovídající typickému lidskému kolísání pozornosti. Komunikaci a fotografii spuštěného systému znázorňují obrázky 4 a 5.



Obrázek 4: Schéma zapojení RC + komunikace s DMS

Obrázek 5: Fotografie spuštěného systému, videa jsou promítána na obrazovku počítače

Vzhledem k tomu, že **uživatelská rozhraní jednotlivých platforem se zásadně liší**, byly vyvinuty dva platformově specifické skripty – jeden pro Instagram Reels (primární cíl auditu) a druhý pro TikTok. Skripty sdílejí společnou kostru (inicializace session, komunikace s DMS, logování provedených akcí), ale liší se v selektorech UI prvků, navigační logice a ve způsobu lokalizace video ID v paměti procesu, neboť každá aplikace má vlastní strukturu pohledové hierarchie i vlastní vnitřní datové struktury. Tato architektura zároveň znamená, že systém je nutné průběžně udržovat: každá větší aktualizace cílové aplikace, která mění strukturu UI nebo paměťové rozložení, vyžaduje revizi a aktualizaci příslušného skriptu. Tato údržba spolu s revalidací paměťového rozložení po aktualizacích aplikací představuje jeden z hlavních dlouhodobých implementačních nákladů systému.

Každá vykonaná akce je společně s metadaty videa logována lokálně i odeslána zpět na stranu DMS, kde je přiřazena ke konkrétní auditní session a uložena do databáze pro pozdější analýzu.

5.3 Co řešení přináší

Realizovaný Remote Control umožňuje provádět audit doporučovacích feedů Instagram Reels a TikTok souběžně na čtyřech zařízeních z jediné pracovní stanice, přičemž každé zařízení může vystupovat pod jiným syntetickým profilem a může být přiřazeno do testovací nebo kontrolní skupiny. Architektonické rozdělení na společnou kostru a platformově specifické skripty umožňuje rozšířit systém o další platformy (např. YouTube Shorts) doplněním samostatného skriptu, aniž by bylo nutné zasahovat do komunikační vrstvy s DMS nebo do logovacího mechanismu.

5.4 Decision Maker Systém (DMS)

Cílem subsystému DMS je v reálném čase rozhodovat o tom, jak má syntetický uživatel zareagovat na konkrétní video zobrazené v doporučovacím feedu. Vstupem je multimediální obsah videa spolu s metadaty, výstupem je strukturovaná sekvence akcí, která bude na zařízení vykonána. Klíčovým požadavkem je **věrohodnost a interpretovatelnost** – rozhodnutí musí věrně odrážet preference profilu definovaného auditorem a zároveň musí být zpětně auditovatelné, aby regulátoři i další výzkumníci mohli přesně reprodukovat a verifikovat každou jednotlivou vykonanou akci. Neprůhledný end-to-end neuronový rozhodovací model by tuto verifikaci znemožnil, a ironicky by tak reprodukoval právě *black-box* problém.

Sekundárním požadavkem je rychlost odezvy v jednotkách sekund, což je horní hranice toho, co je při auditu únosné: pomalejší odezva by zkreslovala chování doporučovacího algoritmu, který explicitně používá délku zobrazení videa jako jeden ze signálů.

5.5 Realizace DMS

DMS je realizován jako webová aplikace s backendem v Pythonu nad frameworkem **FastAPI** a frontendem v **Next.js**. Backend tvoří jádro systému: zpřístupňuje veřejný API endpoint pro výpočet doporučené uživatelské interakce a interní endpointy pro autentizaci, správu agentů a správu auditních session. Vstupy jsou validovány prostřednictvím **Pydantic** modelů, které zajišťují typovou bezpečnost a automatickou generaci OpenAPI dokumentace. Autentizace je implementována přes **JWT tokeny**: při prvním spuštění je administrátor vyzván k vytvoření silného hesla, které následně slouží pro přístup do systému. Pro každou auditní session systém generuje unikátní video_id, jenž slouží jako autentizační token volání prediktivního API a po ukončení session je zneplatněn, čímž je zajištěno, že k volání nemůže přistupovat neautorizovaná strana.

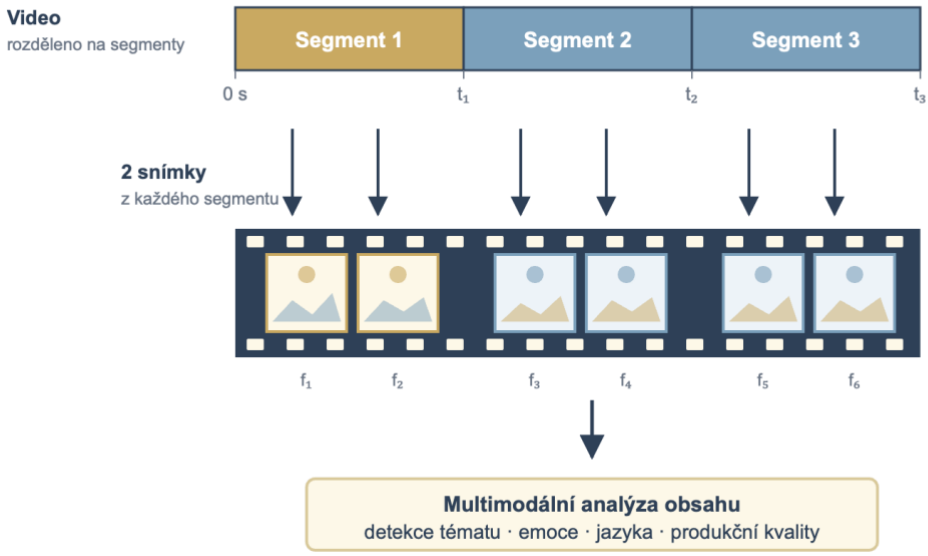
Frontend v Next.js představuje administrační dashboard, ve kterém admin před auditem provádí celou jeho konfiguraci. Jádrem rozhraní je **správa agentů** (CRUD operace) – admin vybere cílovou platformu, definuje strukturovaný uživatelský profil obsahující preference (témata zájmu, preferované emoce, jazyky, oblíbené autory), zvolí výpočetní verzi prediktoru a získává možnost exportovat nasbíraná auditní data vztahující se ke konkrétnímu agentovi.

Druhým modulem je **správa auditních session**, kde admin vybírá agenta, se kterým bude audit proveden, a zakládá samotnou session.

Vlastní rozhodování o uživatelské interakci probíhá ve čtyřech logických fázích. V první fázi systém ověří validitu session tokenu a Pydantic schémat příchozího požadavku; nevalidní požadavky jsou zamítnuty bez dalšího zpracování.

Ve druhé fázi probíhá **zpracování metadat videa**: s pomocí knihovny yt-dlp je staženo samotné video, jsou klasifikovány jeho povrchové charakteristiky (popularita podle počtu zobrazení a interakcí, autor, délka) a video je rozsegmentováno – místo analýzy celého obsahu jsou analyzovány pouze vybrané části v závislosti na délce. Tento přístup vychází z poznatku, že uživatelé se typicky rozhodují v prvních několika sekundách sledování, a proto není ani nutné, ani časově efektivní analyzovat celé video; pro různé délkové kategorie byly definovány specifické strategie segmentace s důrazem na začátek videa a přechody mezi sekvencemi.

Ve třetí fázi probíhá **multimodální analýza obsahu** – z vybraných segmentů jsou pomocí knihovny **OpenCV** extrahovány klíčové snímky, na kterých specializované AI modely detekují téma, dominantní emoci, jazykovou stopu a technickou produkční kvalitu, paralelně je z audio stopy přes **ShazamAPI** rozpoznávána hudba na pozadí. Výstupem této fáze je strukturovaný vektor obsahující všechny relevantní vlastnosti videa v normalizované podobě. Diagram segmentace znázorňuje obrázek 6.



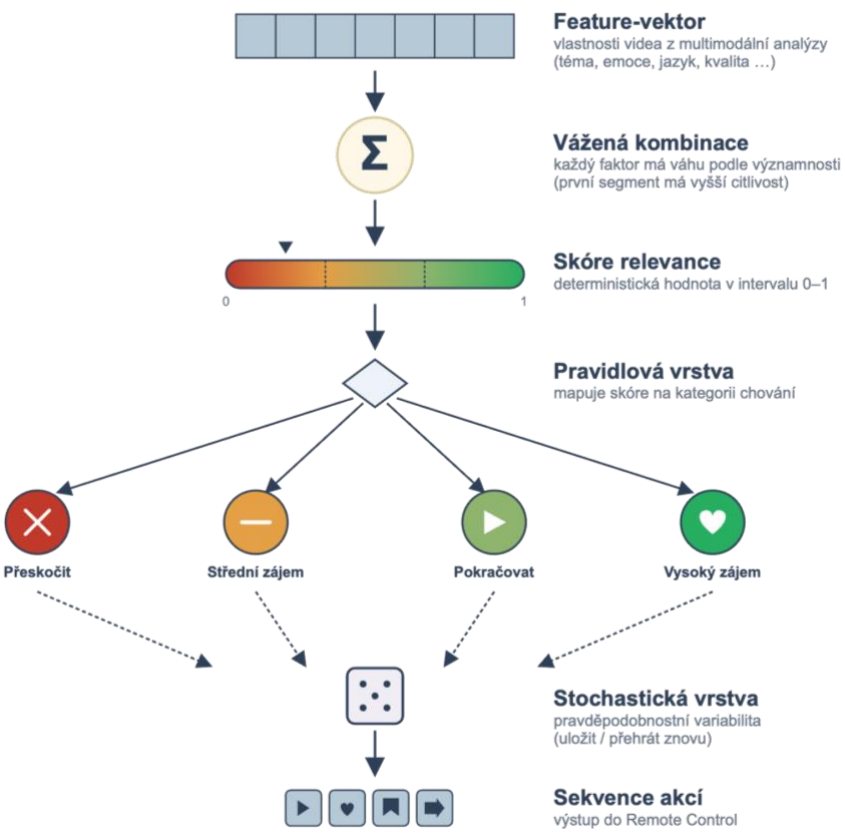
Obrázek 6: Extrakce snímků

Ve čtvrté, rozhodovací fázi vstupuje vektor do **Decision Enginu**, jehož detail tvoří jádro celého subsystému. Decision Engine je hybridní rozhodovací systém kombinující vážené skórování a pravidlové rozhodování nad jednotlivými segmenty videa. Nejprve provádí deterministický výpočet relevance v podobě agregovaného skóre v intervalu <0, 1> které vzniká váženou kombinací faktorů jako emocionální shoda profilu s detekovanou emocí videa, jazyková kompatibilita, tematická relevance, vizuální dynamika, produkční kvalita, úspěšnost detekce hudby, signály retence a bonifikace za vysoký engagement nebo přítomnost oblíbeného autora. Váhy jednotlivých faktorů se liší podle toho, zda je vyhodnocován první segment videa, nebo některý z následujících – první segment má vyšší citlivost na obsahové a emoční signály, neboť právě v něm se uživatel typicky rozhoduje, zda video sledovat dál.

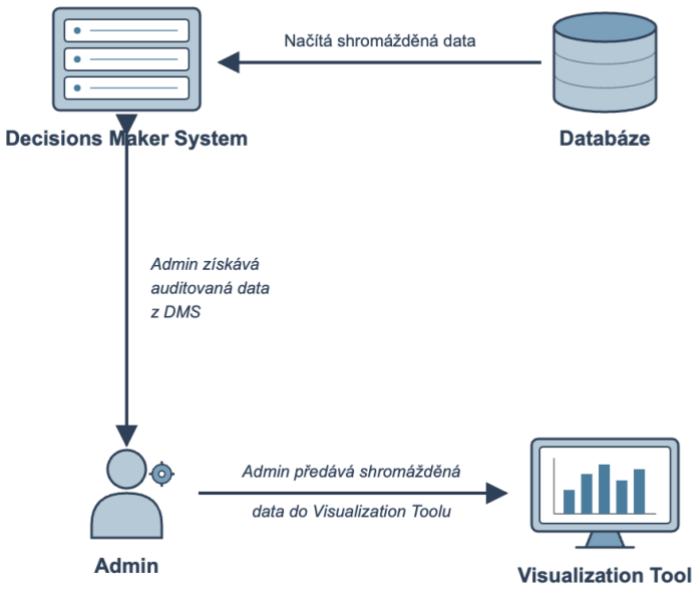
Na vypočtené skóre je následně aplikována **pravidlová rozhodovací vrstva**, která hodnotu skóre mapuje do kategorií chování: vysoký zájem (vede k uložení nebo opětovnému přehrání), pokračování ve sledování, střední zájem (pasivní dokoukání bez další interakce), přeskočení. Tato vrstva zároveň zohledňuje kontext segmentu v rámci videa – penalizuje skóre při postupující divácké únavě v pozdějších segmentech a aplikuje speciální logiku pro poslední segment, kde může dojít k dokončení sledování nebo ke zvýšené interakci typu komentář či sdílení. Nad tímto deterministickým základem leží vrstva, která u vybraných stavů zavádí pravděpodobnostní rozhodování pro akce typu uložení nebo opětovné přehrání. Tím je

dosaženo přirozené variability výstupu i mezi vícenásobnými auditními session se stejným profilem nad stejným videem, aniž by byla narušena konzistence relevančního skóre. Celý rozhodující proces je pro zjednodušení znázorněn obrázkem 7. Před odesláním zpět na Remote Control je výsledná sekvence akcí ještě **validována, deduplikována a kontextově upravena**. Validace ověřuje, že akce odpovídají schopnostem cílové platformy, deduplikace odstraňuje vícenásobné výskyty stejné akce v rámci jedné odpovědi a kontextová úprava zarovnává délku případného pokračování ve sledování ke zbývajícímu času videa, aby agent nepokračoval déle, než video vůbec trvá.

Souběžně s odpovědí pro RC zapisuje DMS plnou stopu rozhodování (vstupní vektor, vypočtené skóre, prošlou větev pravidlové vrstvy, výsledné akce) do databáze. Tato stopa je klíčová pro pozdější analýzu ve **Visualization Toolu** a zároveň zajišťuje zpětnou auditovatelnost rozhodnutí, viz obrázek 8



Obrázek 7: Schéma celého rozhodovacího procesu



Obrázek 8: Schéma datového toku směrem do VT

5.6 Co řešení přináší

Decisions Maker System poskytuje **deterministický a interpretovatelný** rozhodovací mechanismus, který lze pro libovolný produkovaný výstup zpětně rozložit na konkrétní vstupní signály a konkrétní pravidla, jež k danému závěru vedla. Tato vlastnost je v kontextu auditní práce zásadní, neboť respektuje princip, který práce kritizuje u samotných platforem: rozhodovací logika je transparentní a verifikovatelná. Rozdělení do tří vrstev (vstupní validace, multimodální analýza, Decision Engine) navíc umožňuje nezávisle rozšiřovat či nahrazovat jednotlivé komponenty – například přidat do analytické vrstvy nový AI model nebo upravit váhy v Decision Enginu – bez nutnosti zasahovat do zbylých částí systému.

5.7 Visualization Tool (VT)

Visualization Tool tvoří závěrečnou vrstvu auditního systému: na vstupu přijímá CSV export auditních záznamů z Decisions Maker Systemu a na výstupu produkuje sadu vizualizací doplněnou souhrnnou tabulkou. Přímými příjemci výstupů jsou dvě odlišné skupiny – výzkumníci provádějící hlubší analýzu auditních dat a regulátoři či širší veřejnost, kteří potřebují rychlé a interpretovatelné shrnutí. Tato dvojí cílová skupina přímo formuje tři návrhové principy nástroje: **konzistenci** (jednotný vizuální jazyk umožňující porovnávat výsledky napříč tématy a session), **úplnost pohledu** (současné zobrazování relativních i absolutních hodnot, neboť nevyvážené skupiny vedou při samostatném zobrazení procent k systematické dezinterpretaci) a **srozumitelnost** (popisované osy, anotované hodnoty a typ grafu odpovídající otázce, na kterou daný graf odpovídá).

5.8 Realizace VT

Nástroj je realizován jako Jupyter notebook v jazyce Python, postavený na knihovnách **Pandas** (datová transformace), **NumPy** (numerické výpočty), **Matplotlib** (vykreslovací jádro) a **Seaborn** (vyšší stylová vrstva nad Matplotlibem). Volba notebookového prostředí byla vědomá: každá vizualizace je obklopena markdown buňkou popisující analytickou otázku, na kterou graf odpovídá, a interpretací výstupu, takže notebook je čitelný i bez znalosti zdrojového kódu. Tento přístup zároveň usnadňuje využití nástroje nezávislými výzkumnými týmy, které mohou notebook spustit nad vlastním auditním datasetem bez nutnosti nasazovat dedikovanou webovou aplikaci.

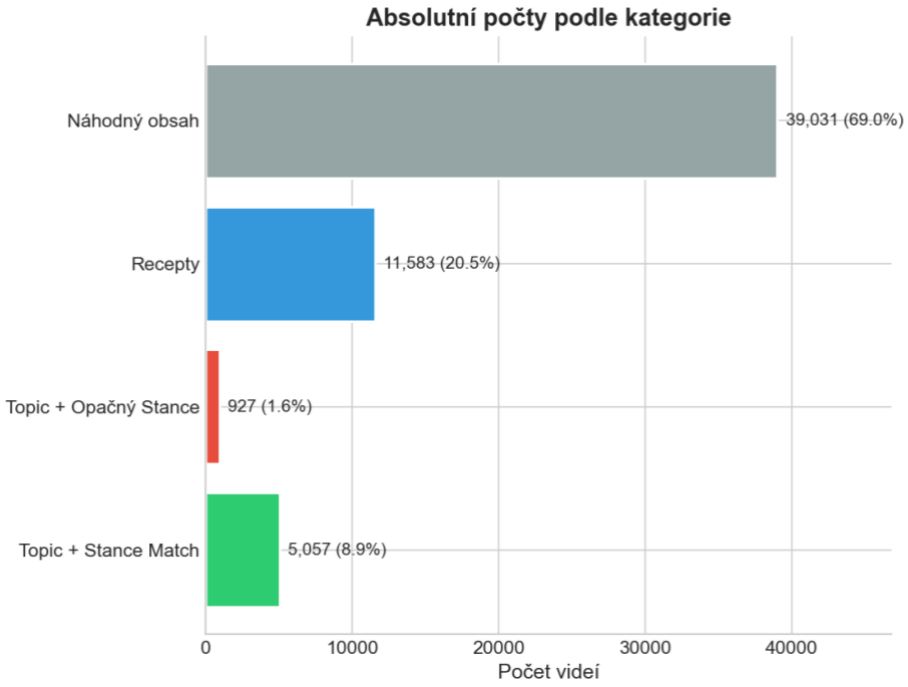
Notebook je členěn do šesti tematických sekcí odpovídajících analytickým vrstvám auditu – globální přehled datasetu, analýza algoritmické bubliny, analýza uživatelských interakcí, přesnost klasifikace, mezitematické srovnání a demografická analýza –, přičemž každá sekce obsahuje sadu samostatných buněk řešících jednu konkrétní analytickou otázku. Celkem nástroj produkuje čtrnáct grafů a souhrnnou tabulku ve formátu CSV, kterou lze přímo zařadit do auditní zprávy.

Vizuální konzistence napříč všemi výstupy je zajištěna centrálně definovanou barevnou paletou a **globálním nastavením parametrů knihovny Matplotlib** v inicializační buňce notebooku. Změna stylu – barev, fontu, velikosti popisků, formátu os – proto vyžaduje úpravu na jediném místě a nevyžaduje zásahy do jednotlivých vykreslovacích buněk. Tento přístup je důležitý pro

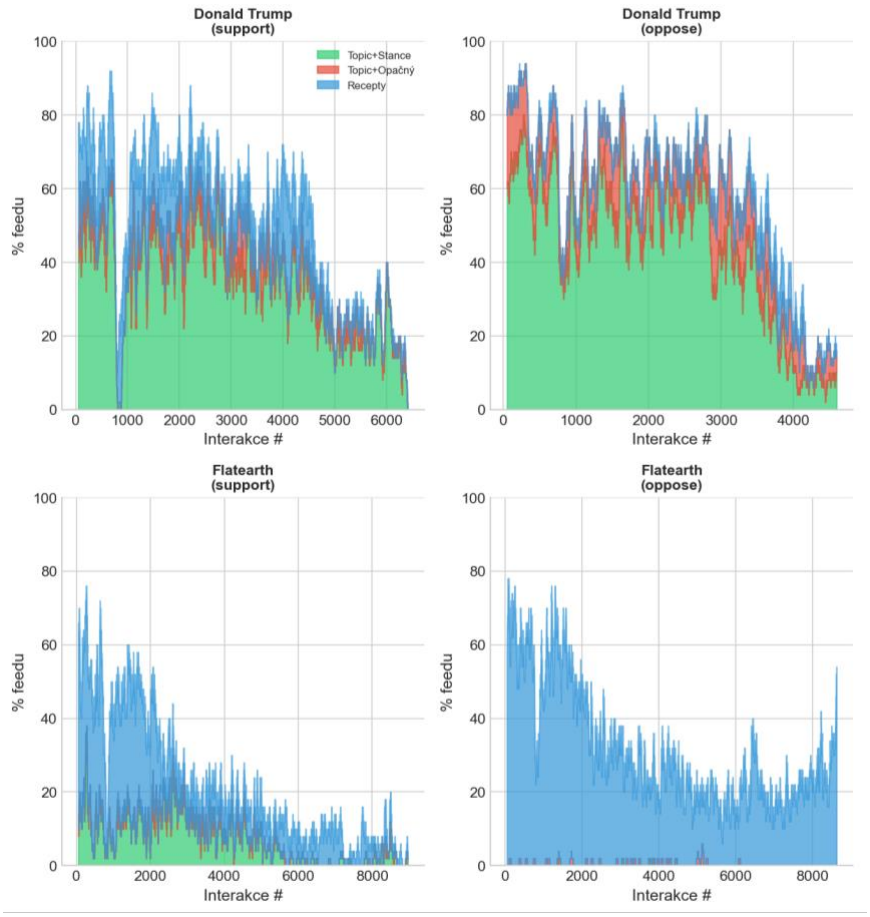
využití výstupů přímo v prezentacích a auditních zprávách, kde by nekonzistentně stylované grafy oslabovaly důvěryhodnost dokumentu.

Pro každou ze čtyř hlavních analytických úloh nástroje byl typ grafu zvolen podle povahy zobrazovaných dat a podle otázky, na kterou má graf odpovídat. Mapování úloh na konkrétní vizualizační metody včetně stručného odůvodnění volby každé z nich znázorňují obrázky 9 a 10

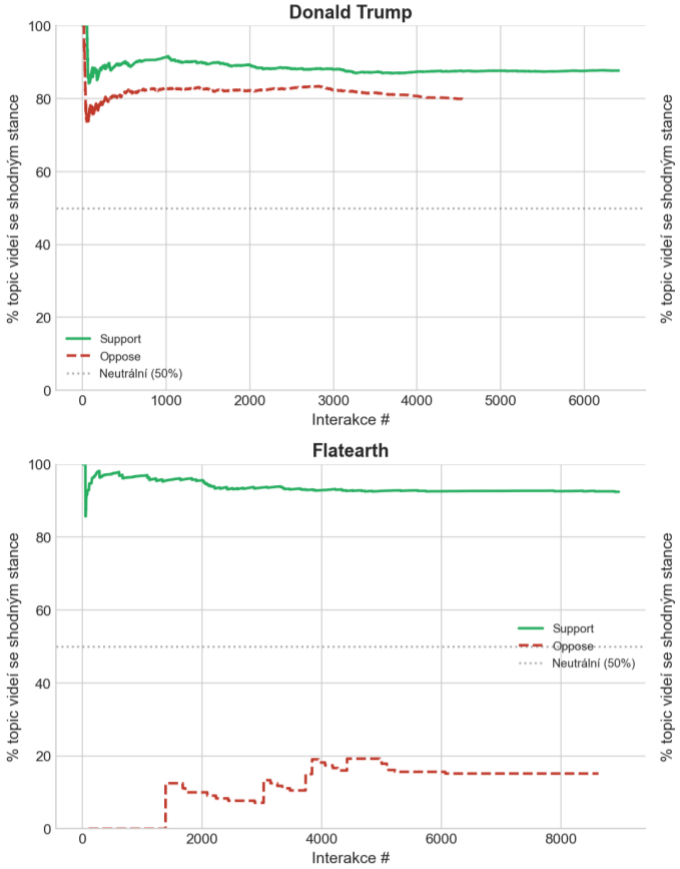
Klíčovým analytickým rozhodnutím bylo zavedení vlastní metriky – **Bubble Indexu** –, která v nástroji slouží pro kvantifikaci efektu algoritmického zužování feedu. Standardní přístup, který sleduje podíl videí potvrzujících postoj uživatele v celém feedu, trpí kombinací dvou efektů: tematické personalizace (algoritmus zužuje feed na témata, jež uživatele zajímají) a jednostranné zaujatosti (algoritmus v rámci daného tématu preferuje konkrétní postoj). Bubble Index tyto efekty odděluje tak, že podíl videí potvrzujících postoj uživatele počítá výhradně ze skupiny videí se shodujícím se tématem. Hodnota indexu blízká 0,5 indikuje, že algoritmus v rámci tématu vystavuje uživatele oběma stranám argumentu vyrovnaně; hodnota výrazně se blížící 0 nebo 1 indikuje, že algoritmus uzavírá uživatele do jedné postojové bubliny. Bubble Index výstup lze vidět na obrázku 11.



Obrázek 9: Globální distribuce videí z první studie KInITu



Obrázek 10: Tzv. "Sliding window" algoritmus ukazující vývoj zobrazovaných videí v čase



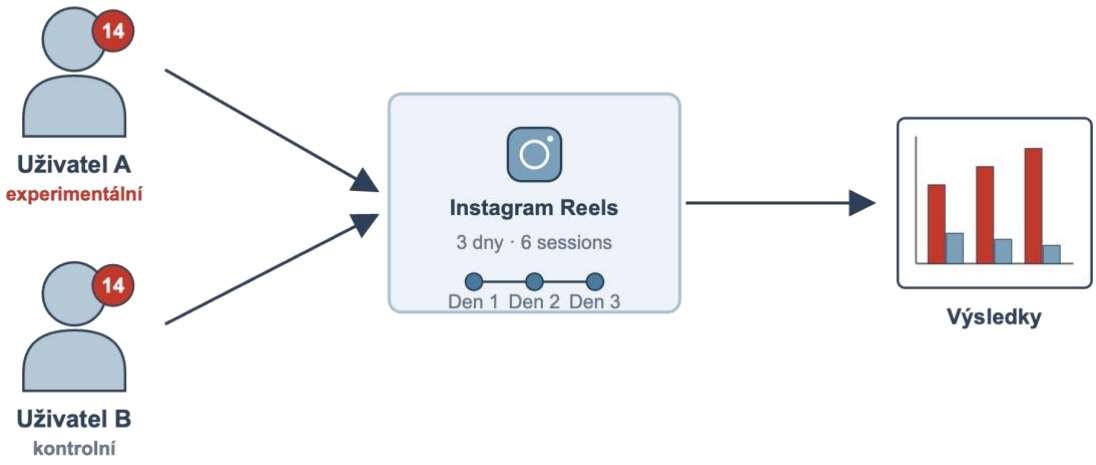
Obrázek 11: Bubble index grafy topiců americké politiky a ploché země

5.9 Co řešení přináší

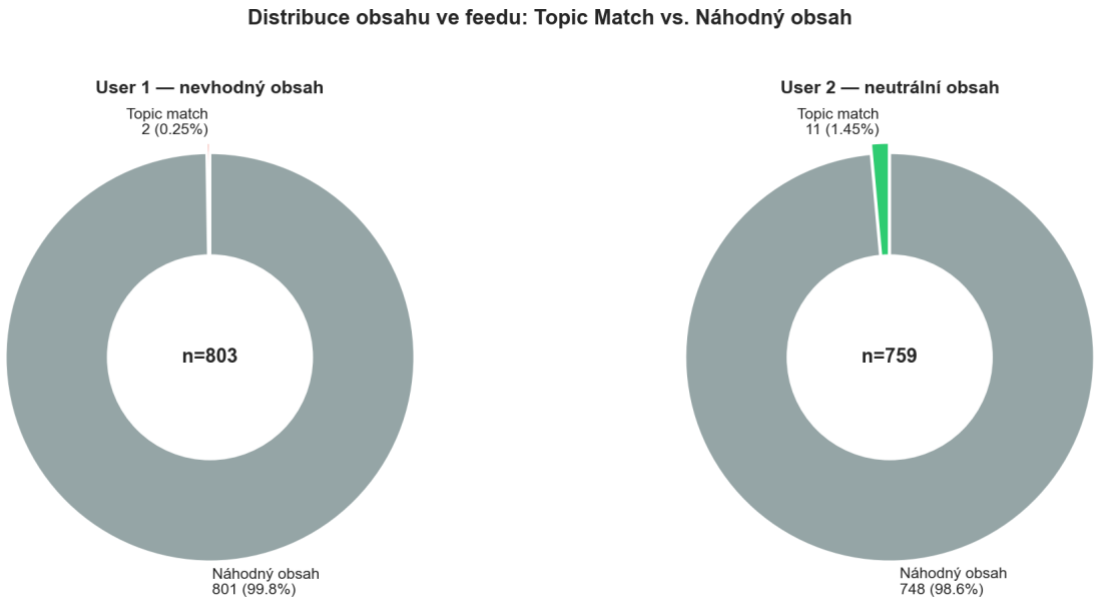
Visualization Tool transformuje auditní data z DMS do interpretovatelné podoby použitelné jak ve vědecké analýze, tak v auditní zprávě určené regulátorovi. Notebooková forma umožňuje cizím výzkumným týmům pustit nástroj nad vlastními daty bez nutnosti nasazovat infrastrukturu, centrální stylová vrstva zajišťuje, že výstupy jdou bez další úpravy přímo do prezentace, a Bubble Index nabízí jednu z mála prakticky použitelných metrik algoritmické bubliny, která **odděluje tematickou a postojovou personalizaci** a poskytuje tak interpretačně ostřejší obraz, než jaký umožňují standardní agregátní metriky.

6 PILOTNÍ AUDIT

Nástroj je ověřen na reálném datasetu získaném z auditu Instagram Reels metodou sock-puppet, který zahrnuje detekci rizikových témat (gambling, zábava pro dospělé, černý humor, obsah s kriminálním kontextem atd.) ve feedu nezletilých uživatelů. Audit slouží primárně pro test celého systému. Časové a technické omezení nám nedovolují audit provést v plné míře. Pro audit byly zvoleny 2 účty, oba ve věku 14 let, kdy u jednoho z nich byl explicitně vyhledán škodlivý obsah. U druhého z nich byla pro klasifikaci topiců vybrána témata zdravého životního stylu, mezi nimi byly například: „renewable_energy, travel, healthy_lifestyle“. Audit je zjednodušeně popsán obrázkem číslo 12.



Obrázek 12: Schéma testovacího auditu



Obrázek 13: Globální distribuce napříč sessions

Mezi filtrovaným vzorkem dat zatím nebyl na platformě Instagram reels nalezen žádný vzorec promování rizikového obsahu u uživatelů, kteří nevhodný obsah vyhledali explicitně. 2 videa v rámci výskytu byla prediktorem označena jako zavádějící. Po zpětném prohlédnutí ale videa netvoří závažný obsah. Viz obrázek 13.

- https://www.instagram.com/reel/DRyFmCoDSTg/ - prediktorem vyhodnoceno jako „gambling“ (Video ukazující trik s mincí ve stylu skořápek)
- https://www.instagram.com/reel/DTAxyrej-ov/ - prediktorem vyhodnoceno jako „taboo_topics“ (Video s dítětem cca 1 rok starým, které je plácáno po tváři)

Ačkoliv Instagram vyhledání rizikového obsahu umožňuje, upozorňuje při tom na vyhledání takového obsahu přímo v rozhraní aplikace, ať už cenzurou příspěvků, nebo upozorněním v podobě pop-up okna. Nezletilým uživatelům se však obsah ukáže velice filtrovaný, nebo žádný. Řadí se sem obsah primárně se sexuálním kontextem, rasistickým podtextem nebo tzv. „gore“ videa a další.

6.1 Limity

Vyvinutý systém přináší funkční základ pro automatizované algoritmické audity, jeho aplikovatelnost je však omezena v několika ohledech.

Agent je závislý na stabilním přístupu k rootnutému Android zařízení a na konkrétní struktuře uživatelského rozhraní aplikace Instagram – jakákoli změna rozložení prvků na obrazovce může vyžadovat aktualizaci scriptu. V rámci auditu například způsobovalo komplikace pop-up okno, které upozorňovalo na příliš dlouhou aktivitu na Instagramu. Tato okna jsou ale pro nezletilé fixně vymezena a v novějších verzích již nejde vypnout jejich výskyt.

Prediktivní systém dosahuje průměrné doby odezvy 5-30 sekund, což pro auditní účely postačuje, nicméně při analýze velmi krátkých videí nebo při výpadku externích služeb (Shazam API) může být přesnost predikce snížena.

Vizualizační nástroj předpokládá správnou klasifikaci tématu a postoje ve vstupních datech – chyby v předřazené klasifikaci se přímo promítají do výsledků auditu, aniž by je nástroj sám zachytil. Zároveň s každou auditní otázkou je třeba přímého zásahu do vizualizačních skriptů, neboť není možné pokrýt kompletní spektrum toho, co je možné auditovat.

Celý systém byl ověřen na omezeném vzorku účtů a platforem, takže zobecnění výsledků na širší populaci uživatelů nebo jiné platformy vyžaduje další validaci. V neposlední řadě sock-puppet metodologie zpravidla porušuje podmínky použití auditovaných platforem, což může v budoucnu vést k technickým protiopatřením ze strany provozovatelů.

7 ZÁVĚR

Tato práce si kladla za cíl navrhnout systém pro automatizované algoritmické audity, který překonává metodologické nedostatky identifikované v první vlně DSA auditů – zejména absenci standardizace a nesrovnatelnost výsledků mezi platformami. Navržený třísložkový systém – Remote Control, Decisions Maker System a vizualizační systém – cíl naplňuje: umožňuje opakovatelné a strukturované audity bez nutnosti přístupu k interním datům platforem. Systém tak představuje praktický nástroj využitelný jak výzkumnou komunitou, tak regulátory usilujícími o nezávislé ověření chování doporučovacích algoritmů. Dalším přirozeným krokem by bylo rozšíření systému o podporu dalších platforem, implementaci statistického testování hypotéz ve vizualizačním modulu a dlouhodobé testování odolnosti agenta vůči změnám uživatelského rozhraní.

7.1 Význam práce a další možná pokračování výzkumu:

Tato práce otevírá možnosti dalšího výzkumu a pokroku, a to hned v několika směrech. Na práci mohou navázat další práce v oblasti:

- **IT** – tvorba obdobných systémů pro další sociální sítě,
- **sociologie, politologie, psychologie** – nástroj dává možnost vědcům z těchto oborů relativně přesně měřit chování algoritmů na jejich vlastních experimentech a z jejich chování potom vyvozovat důsledky pro chování jednotlivců a skupin.
- **právo** – práce dává základní nástroj pro stanovení a kontrolu veličin a stanovení přípustných hodnot (např.: min. poměr nabízených příspěvků, které nejsou v souladu s názorem uživatele), které mohou regulační orgány stanovit pro poskytovatele sociálních sítí a zároveň kontrolovat jejich dodržování.

Zde jsou odkazy na GitHub ke zdrojovým kódům celého systému:

- https://github.com/MartinPytlik/Instagram_reels_agent
- https://github.com/erikmacak/user-interaction-predictor
- https://github.com/Tvarohaacek/Jupyter3

8 SEZNAM OBRÁZKŮ

9 SEZNAM POUŽITÉ LITERATURY

ADA LOVELACE INSTITUTE, 2021. *Technical methods for regulatory inspection of algorithmic systems*. Online. Neuvedeno: Ada Lovelace Institute [vid. 2026-03-19]. Dostupné z: https://www.adalovelaceinstitute.org/report/technical-methods-regulatory-inspection/

AGYAPONG-OPOKU, Nadine; Felix AGYAPONG-OPOKU a Andrew J. GREENSHAW, 2025. Effects of Social Media Use on Youth and Adolescent Mental Health: A Scoping Review of Reviews. *Behavioral Sciences*. Online. **15**(5), 574. ISSN 2076-328X. Dostupné z: doi:10.3390/bs15050574

ALLEN, Asha, 2025. CDT Europe Responds to the Draft Guidelines on the Online Protection of Minors. *Center for Democracy and Technology*. Online. [vid. 2026-03-19]. Dostupné z: https://cdt.org/insights/cdt-europe-responds-to-the-draft-guidelines-on-the-online-protection-of-minors/

BOUCHAUD, Paul a Pedro RAMACIOTTI, 2024. Auditing the audits: evaluating methodologies for social media recommender system audits. *Applied Network Science*. Online. **9**(1), 59. ISSN 2364-8228. Dostupné z: doi:10.1007/s41109-024-00668-6

CADE PROJECT, 2025. EU guidelines on keeping children safe online under the Digital Services Act. *CADE – Civil Society Alliances for Digital Empowerment*. Online. [vid. 2026-03-19]. Dostupné z: https://cadeproject.org/updates/eu-guidelines-on-keeping-children-safe-online-under-the-digital-services-act/

DE, Debasmita; Mazen EL JAMAL; Eda AYDEMIR a Anika KHERA, 2025. Social Media Algorithms and Teen Addiction: Neurophysiological Impact and Ethical Considerations. *Cureus*. Online. [vid. 2026-03-19]. ISSN 2168-8184. Dostupné z: doi:10.7759/cureus.77145

DSA OBSERVATORY, 2024a. *DSA risk assessment reports: A guide to the first rollout and what's next – DSA Observatory*. Online. [vid. 2026-03-19]. Dostupné z: https://dsa-observatory.eu/2024/12/09/dsa-risk-assessment-reports-are-in-a-guide-to-the-first-rollout-and-whats-next/

DSA OBSERVATORY, 2024b. *The Regulation of Recommender Systems Under the DSA: A Transition from Default to Multiple and Dynamic Controls? - DSA Observatory*. Online. [vid. 2026-03-21]. Dostupné z: https://dsa-observatory.eu/2024/11/22/the-regulation-of-recommender-systems-under-the-dsa-a-transition-from-default-to-multiple-and-dynamic-controls/

EU, 2025. *Digital Services Act (DSA) | Updates, Compliance, Training*. Online [vid. 2026-03-19]. Dostupné z: https://www.eu-digital-services-act.com/

EUROPEAN COMMISSION, 2025a. Better Internet for kids | Shaping Europe's digital future. *European Commission*. Online [vid. 2026-03-19]. Dostupné z: https://digital-strategy.ec.europa.eu/en/factpages/better-internet-kids

EUROPEAN COMMISSION, 2025b. *Digital Services Act: keeping us safe online - European Commission*. Online [vid. 2026-03-19]. Dostupné z: https://commission.europa.eu/news-and-media/news/digital-services-act-keeping-us-safe-online-2025-09-22_en

EUROPEAN PARLIAMENT, 2024. A guide to the Digital Services Act, the EU's law to rein in Big Tech. *AlgorithmWatch*. Online [vid. 2026-03-19]. Dostupné z: https://algorithmwatch.org/en/dsa-explained/

EUROPEAN PARLIAMENT, 2026. The impact of the Digital Services Act on digital platforms | Shaping Europe's digital future. *European Parliament*. Online [vid. 2026-03-19]. Dostupné z: https://digital-strategy.ec.europa.eu/en/policies/dsa-impact-platforms

GOODMAN, Emma, 2025. Understanding the effects of social media on children - Media@LSE. *Media@LSE - Promoting critical research into the vital role of media and communications in contemporary society*. Online. [vid. 2026-03-19]. Dostupné z: https://blogs.lse.ac.uk/medialse/2025/09/11/understanding-the-effects-of-social-media-on-children/

HOGAN LOVELLS, 2025. The long-awaited EU Guidelines on Article 28(1) DSA: What online platforms must know. *www.hoganlovells.com*. Online [vid. 2026-03-19]. Dostupné z: https://www.hoganlovells.com/en/publications/the-longawaited-eu-guidelines-on-article-281-dsa-what-online-platforms-must-know

IYER, Ram, 2025. EC finds Meta and TikTok breached transparency rules under DSA. *TechCrunch*. Online. [vid. 2026-03-19]. Dostupné z: https://techcrunch.com/2025/10/24/ec-finds-meta-and-tiktok-breached-transparency-rules-under-dsa/

JALLI, Nuurrianti, 2024. Holding Social Media Companies Accountable for Enabling Hate and Disinformation. (2024).

JOINT RESEARCH CENTRE, 2025. *Why are children and adolescents vulnerable to social media? - Joint Research Centre*. Online [vid. 2026-03-19]. Dostupné z: https://joint-research-centre.ec.europa.eu/jrc-explains/why-are-children-and-adolescents-vulnerable-social-media_en

KOZYREVA, Anastasia; Stefan M. HERZOG; Stephan LEWANDOWSKY; Ralph HERTWIG; Philipp LORENZ-SPREEN; Mark LEISER a Jason REIFLER, 2022. Resolving content moderation dilemmas between free speech and harmful misinformation. *Proceedings of the National Academy of Sciences of the United States of America*. Online. **120**(7), e2210666120. ISSN 0027-8424. Dostupné z: doi:10.1073/pnas.2210666120

KRAFFT, Tobias D.; Marc P. HAUER a Katharina ZWEIG, 2024. Black-Box Testing and Auditing of Bias in ADM Systems. *Minds and Machines*. Online. **34**(2), 15. ISSN 1572-8641. Dostupné z: doi:10.1007/s11023-024-09666-0

LAHTI, Henri; Marja KOKKONEN; Lauri HIETAJÄRVI; Nelli LYYRA a Leena PAAKKARI, 2024. Social media threats and health among adolescents: evidence from the health behaviour in school-aged children study. *Child and Adolescent Psychiatry and Mental Health*. Online. **18**(1), 62. ISSN 1753-2000. Dostupné z: doi:10.1186/s13034-024-00754-8

MEDIALAWS, 2025. €120 million later: the DSA enters the enforcement phase. *MediaLaws*. Online. [vid. 2026-03-19]. Dostupné z: https://www.medialaws.eu/e120-million-later-the-dsa-enters-the-enforcement-phase/

NAVEA, Alan Friel, Francesco Liberatore, Gorka, 2024. EU Digital Services Act in Full Force. *Privacy World*. Online. [vid. 2026-03-19]. Dostupné z: https://www.privacyworld.blog/2024/02/eu-digital-services-act-in-full-force/

OHNESORGE, Jella, 2025. *Counting without accountability? An analysis of the DSA's transparency reports*. Online. 25. září 2025. Neuvedeno: Zenodo. [vid. 2026-03-19]. Dostupné z: doi:10.5281/ZENODO.17201618

OJEWALE, Victor; Ryan STEED; Briana VECCHIONE; Abeba BIRHANE a Inioluwa Deborah RAJI, 2025. Towards AI Accountability Infrastructure: Gaps and Opportunities in AI Audit Tooling. In: *CHI 2025: CHI Conference on Human Factors in Computing Systems*: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Online. Yokohama Japan: ACM, s. 1–29 [vid. 2026-03-21]. ISBN 979-8-4007-1394-1. Dostupné z: doi:10.1145/3706598.3713301

PARLIAMENT, European, 2024. 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 | Legislative Train Schedule. *European Parliament*. Online [vid. 2026-03-19]. Dostupné z: https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-digital-services-act

ROSS ARGUEDAS, Amy; Craig T. ROBERTSON; Richard FLETCHER a Rasmus Kleis NIELSEN, 2022. *Echo chambers, filter bubbles, and polarisation: a literature review*. Online. Neuvedeno: Reuters Institute for the Study of Journalism [vid. 2026-03-19]. Dostupné z: doi:10.60625/RISJ-ETXJ-7K60

ROY, Ash Johnson, Puja, 2025. *EU Should Improve Transparency in the Digital Services Act*. Online. [vid. 2026-03-19]. Dostupné z: https://itif.org/publications/2025/10/20/eu-should-improve-transparency-in-the-digital-services-act/

SHIN, Donghee a Kulsawasd JITKAJORNWANICH, 2024. How Algorithms Promote Self-Radicalization: Audit of TikTok's Algorithm Using a Reverse Engineering Method. *Social*

Science Computer Review. Online. **42**(4), 1020–1040. ISSN 0894-4393, 1552-8286. Dostupné z: doi:10.1177/08944393231225547

SOLAROVA, Sara; Matúš MESARČÍK; Branislav PECHER a Ivan SRBA, 2026. *Beyond the Checkbox: Strengthening DSA Compliance Through Social Media Algorithmic Auditing*. Online. 26. leden 2026. [vid. 2026-03-19]. Dostupné z: doi:10.1145/3772318.3791774

SRBA, Ivan; Branislav PECHER; Jakub SIMKO; Robert MORO a Maria BIELIKOVA, 2025. Model-based Algorithmic Auditing of Social Media AI Algorithms.

STARKE, Christopher; Ljubiša METIKOŠ; Natali HELBERGER a Claes DE VREESE, 2025. Contesting personalized recommender systems: a cross-country analysis of user preferences. *Information, Communication & Society*. Online. **28**(1), 41–60. ISSN 1369-118X, 1468-4462. Dostupné z: doi:10.1080/1369118X.2024.2363926

TECH POLICY PRESS, 2024. *5 Things to Know about the Digital Services Act's First Risk Assessments and Audits | TechPolicy.Press*. Online [vid. 2026-03-19]. Dostupné z: https://www.techpolicy.press/5-things-to-know-about-the-digital-services-acts-first-risk-assessments-and-audits/

TRUJILLO, Amaury; Tiziano FAGNI a Stefano CRESCI, 2025. The DSA Transparency Database: Auditing Self-reported Moderation Actions by Social Media. *Proceedings of the ACM on Human-Computer Interaction*. Online. **9**(2), 1–28. ISSN 2573-0142. Dostupné z: doi:10.1145/3711085

VCARD, 2024. The Digital Services Act Is Now Fully Applicable and Enforceable. *Steptoe*. Online [vid. 2026-03-19]. Dostupné z: https://www.steptoe.com/en/news-publications/steptechtoe-blog/the-digital-services-act-is-now-fully-applicable-and-enforceable.html

WANG, Jingsong a Shen WANG, 2025. The Emotional Reinforcement Mechanism of and Phased Intervention Strategies for Social Media Addiction. *Behavioral Sciences*. Online. **15**(5), 665. ISSN 2076-328X. Dostupné z: doi:10.3390/bs15050665

YE, Jinyi; Luca LUCERI a Emilio FERRARA, 2025. Auditing Political Exposure Bias: Algorithmic Amplification on Twitter/X During the 2024 U.S. Presidential Election. In: *FAccT '25: The 2025 ACM Conference on Fairness, Accountability, and Transparency*: *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*. Online. Athens Greece: ACM, s. 2349–2362 [vid. 2026-03-19]. ISBN 979-8-4007-1482-5. Dostupné z: doi:10.1145/3715275.3732159

ZHOU, Ren, 2024. Understanding the Impact of TikTok's Recommendation Algorithm on User Engagement. *International Journal of Computer Science and Information Technology*. Online. **3**(2), 201–208. ISSN 3005-7140, 3005-9682. Dostupné z: doi:10.62051/ijcsit.v3n2.24