

**DELTA – Střední škola informatiky a ekonomie, s.r.o.**

Ke Kamenci 151, Pardubice

**Auditování algoritmů sociálních sítí s využitím umělé  
inteligence**

**Autor:** Erik Macák, Martin Pytlík, Michal Tvaroh

**Třída:** 4.A, 4.B

**Studijní obor:** Informační technologie (18-20-M/01)

**Rok:** 2025/2026

# Zadání maturitního projektu z informatických předmětů

Jméno a příjmení:	<i>Erik Macák</i>
Pro školní rok:	<i>2025/2026</i>
Třída:	<i>4. B</i>
Obor:	<i>Informační technologie 18-20-M/01</i>
Téma práce:	<i>Prediktivní systém pro simulaci uživatelských interakcí na sociálních sítích</i>
Vedoucí práce:	<i>Ivan Srba</i>

Způsob zpracování, cíle práce, pokyny k obsahu a rozsahu práce:

## ***Cíl projektu:***

Algoritmy sociálních sítí, jako jsou doporučovací systémy či vyhledávače, dnes výrazně ovlivňují, jaký obsah uživatelé vidí a jak se na internetu chovají. Aby bylo možné jejich fungování lépe pochopit a kontrolovat, vznikají tzv. algoritmické audity – procesy, které zkoumají chování platforem pomocí simulovaných uživatelských interakcí. Současné audity však často pracují s příliš jednoduchými modely uživatelů, které neodrážejí reálné lidské rozhodování. Proto je potřeba vytvářet pokročilejší prediktivní systémy, které dokážou předpovídat skutečné uživatelské reakce a tím zpřesnit výsledky auditů. Cílem tohoto projektu je vytvořit vlastní prediktivní systém (predictor), který dokáže předpovídat uživatelské interakce na sociálních sítích v rámci doporučovacích algoritmů. Systém bude pracovat s modely lidského chování a bude schopen imitovat reálné rozhodování uživatele při interakci s obsahem. Systém bude přijímat identifikátory konkrétního obsahu, zpracuje je prostřednictvím algoritmické a případně AI vrstvy, vyhodnotí vhodnou reakci a pošle odpověď zpět. Projekt zároveň umožňuje hlubší vhled do problematiky AI a auditu algoritmů sociálních sítí s důrazem na etické aspekty a transparentnost automatizovaného rozhodování.

## **Specifikace projektu:**

### **1. Výzkum:**

- Bude proveden výzkum problematiky doporučovacích algoritmů sociálních sítí a existujících přístupů k modelování uživatelského chování.
- Bude analyzován vliv AI metod na predikci lidského rozhodování a jejich použitelnost pro tento projekt.

### **2. Analýza a návrh:**

- Budou definovány požadavky na systém.
- Bude navržena architektura systému.
- Bude navrženo rozhraní pro příjem identifikátorů obsahu a odesílání výsledných reakcí.
- Bude vytvořen návrh toků dat v systému.

### **3. Vývoj systému:**

- Bude implementována funkcionální pro příjem požadavků, zpracování obsahu a generování simulovaných interakcí.
- Bude implementována logika predikce uživatelských reakcí na základě relevance.
- Bude zajištěna modularita kódu pro snadnou údržbu a rozšiřitelnost.

### **4. Testování a dokumentace:**

- Systém projde jednotkovými testy, integračními testy a testy funkčnosti predikcí.
- Bude provedeno testování robustnosti a výkonu systému při zpracování většího množství požadavků.
- Bude vytvořena uživatelská a vývojářská dokumentace, včetně popisu architektury, použitých technologií a způsobu nasazení.

### **5. Kvalita kódu a architektura systému:**

- Kód bude psán s důrazem na principy Clean Code.
- Architektura bude modulární a vícevrstvá, aby byla zajištěna snadná údržba a budoucí rozšiřitelnost.
- Bude dodržováno verzování kódu a standardy pro code review.

### **6. Prezentace a obhajoba:**

- Bude připravena prezentace projektu, která představí hlavní funkce aplikace.
- Bude předvedena ukázka fungování systému.
- Bude připravena obhajoba projektu zahrnující technické detaily, použitou metodiku, zhodnocení výsledků a získaných zkušeností.

### ***Požadované výstupy:***

- Zpracovaná analýza doporučovacíh algoritmů a modelů lidského chování.
- Implementované rozhraní pro příjem vstupních dat a odesílání výsledků.
- Generování a vracení relevantních uživatelských interakcí s ohledem na postavení uživatele a auditní rámec.
- Zpracování obsahu a predikce odpovídajících reakcí.
- Uživatelská a vývojářská dokumentace, včetně popisu architektury, použitých technologií a způsobu nasazení.
- Prezentace projektu a ukázka jeho funkcionality.
- Kompletní zdrojové kódy včetně testů.
- Zajištění funkčnosti celého systému dle definovaných požadavků.

### ***Hodnocení:***

Projekt bude hodnocen na základě následujících kritérií:

- obecné povědomí o problematice doporučovacíh algoritmů,
- implementace a správnost komunikace systému (příjem vstupních dat, odesílání výsledků),
- zpracování obsahu a predikce uživatelských interakcí,
- implementace a celková funkčnost systému,
- úroveň dokumentace a prezentace projektu,
- kreativita a efektivnost řešení.

Stručný časový harmonogram (s daty a konkretizovanými úkoly):

- **Září:** Seznámení s prostředím v KInIT a nastavení týmové spolupráce (workflow, role, nástroje), úvodní výzkum problematiky doporučovacích algoritmů a uživatelského chování, analýza dostupných technologií a návrh technologického stacku.
- **Říjen:** Definice požadavků na systém, návrh architektury aplikace a rozhraní pro příjem a odesílání dat, studium klíčových principů AI pro následnou implementaci prediktivní logiky.
- **Listopad-prosinec:** Samotná implementace včetně zpracování požadavků a základní aplikační logiky, postupné napojení na komponenty pro zpracování obsahu a predikci interakcí, testování dílčích částí.
- **Leden:** Integrace všech komponent do funkčního celku, optimalizace výkonu a efektivity systému, testování z pohledu funkčnosti, robustnosti a bezpečnosti, ladění chyb a stabilizace.
- **Únor–březen:** Vytvoření uživatelské a vývojářské dokumentace, závěrečné testování a formální uzavření vývoje, příprava a nácvik prezentace a obhajoby projektu.

Prohlašuji, že jsem maturitní projekt vypracoval(a) samostatně, výhradně s použitím uvedené literatury.

V Pardubicích 24. 3. 2026

.....

*(vlastnoruční podpis)*

# Zadání maturitního projektu z informatických předmětů

Jméno a příjmení: *Martin Pytlík*

Pro školní rok: *2025/2026*

Třída: *4. A*

Obor: *Informační technologie 18-20-M/01*

Téma práce: *Rozšíření systému pro simulaci uživatelského chování na platformu Instagram Reels*

Vedoucí práce: *Ivan Srba*

Způsob zpracování, cíle práce, pokyny k obsahu a rozsahu práce:

## ***Cíl projektu:***

Cílem tohoto projektu je rozšířit stávající systém pro automatizovanou simulaci uživatelského chování, původně vyvinutý pro platformu TikTok, na mobilní aplikaci Instagram se zaměřením na funkci Reels. Projekt bude demonstrovat možnosti využití nástrojů pro automatizaci (např. UIAutomator 2) při ovládání mobilní aplikace a vytvoří sadu scénářů simulujících běžné uživatelské interakce (scrollování, zastavení u videa, lajkování, apod.). Systém bude sloužit jako rozšiřitelný modul, který bude možné začlenit do širšího výzkumného projektu AI-Auditology.

## **Specifikace projektu:**

### **1. Analýza a návrh:**

- Seznámit se s již existujícím řešením pro TikTok.
- Identifikovat rozdíly mezi prostředím TikToku a Instagramu (UI, interakce uživatele).
- Navrhnout přizpůsobení simulátoru chování pro prostředí Instagram Reels.

### **2. Výzkum a návrh technologií:**

- Ověřit vhodnost stávajícího technologického stacku pro Instagram.
- Vybrat nástroj pro automatizaci ovládání (UIAutomator 2 nebo alternativy).
- Navrhnout strukturu nových scénářů a jejich integraci do existujícího systému.

### **3. Implementace rozšíření:**

- Implementovat základní funkce: otevření aplikace, přechod do Reels, scrollování, apod.
- Přidat možnosti interakcí: zastavení u vybraného obsahu, lajkování.
- Doplnit variabilní chování (náhodné pauzy, různá délka sledování videa).

### **4. Testování a dokumentace:**

- Ověřit funkčnost rozšíření na reálném zařízení.
- Testovat stabilitu při dlouhodobém běhu scénářů.
- Dokumentovat zdrojový kód, popsat implementaci a způsob použití.
- Stručně shrnout možnosti využití výsledku v rámci analýzy doporučovacích systémů (navazující výzkumný projekt).

## **Hodnocení:**

Projekt bude hodnocen na základě následujících kritérií:

- kvality a úplnosti implementovaného rozšíření,
- funkčnosti a spolehlivosti simulovaného chování na Instagram Reels,
- úrovně dokumentace a přehlednosti kódu,
- schopnosti integrovat řešení do širšího projektu,
- prezentace výsledků a obhajoby návrhových rozhodnutí.

Stručný časový harmonogram (s daty a konkretizovanými úkoly):

- **Září:** Studium existujícího řešení pro TikTok, analýza prostředí Instagramu
- **Říjen:** Návrh scénářů chování pro Instagram Reels, volba technologií
- **Listopad:** Implementace základních funkcí (spuštění, scrollování, přechod do Reels)
- **Prosinec-leden:** Rozšíření o interakce a variabilní chování
- **Únor:** Rozšíření o interakce a variabilní chování
- **Březen:** Dokumentace a příprava prezentace

Prohlašuji, že jsem maturitní projekt vypracoval(a) samostatně, výhradně s použitím uvedené literatury.

V Pardubicích 24. 3. 2026

.....  
(vlastnoruční podpis)

# Zadání maturitního projektu z informatických předmětů

Jméno a příjmení:	<i>Michal Tvaroh</i>
Pro školní rok:	<i>2025/2026</i>
Téma práce:	<i>Návrh datové struktury a zpracování aplikace pro vizualizaci záznamů z projektu AI-Auditology</i>
Vedoucí práce:	<i>Ivan Srba</i>

Způsob zpracování, cíle práce, pokyny k obsahu a rozsahu práce:

## ***Cíl projektu:***

Cílem tohoto projektu je navrhnout vhodnou datovou strukturu pro ukládání záznamů (logů) z auditů v aplikaci TikTok. Výstupem bude aplikace, která uživateli umožní vizualizovat záznamy pomocí vhodné grafické reprezentace. Zároveň bude výstupem odůvodněná volba řešení, která umožní efektivní filtrování a vyhledávání v datech a poskytne uživateli nástroje pro interaktivní práci s logy. Výsledný nástroj by měl být vhodný pro analytické účely a umožnit reprezentaci logů v potřebných zobrazeních. Konkrétní zadání řeší problém přehlednosti, orientace v projektu a interpretace jeho výsledků. Celý projekt AI-Auditology se zaměřuje na vývoj metod pro automatizovaný audit algoritmů sociálních sítí, zejména doporučovacích systémů. Cílem je odhalovat a hodnotit šíření škodlivého obsahu a dezinformací pomocí modelově řízeného auditu, který simuluje chování uživatelů. Výsledkem bude nástroj umožňující nezávislý dohled nad algoritmy velkých platforem v souladu s evropskými hodnotami.

## **Specifikace projektu:**

### **1. Analýza a výzkum:**

- Bude konzultováno vhodné datové uspořádání logů, které bude odpovídat požadavkům projektu a aplikace.
- Budou analyzovány existující technologie, způsoby řešení a přístupy k práci s logy a vědeckými daty.
- Bude navrženo a odůvodněno technické řešení, jazyk, jeho knihovny, platformy a způsobu řešení a užití pro výzkum.

### **2. Rozvržení aplikace:**

- Bude navržen systém, ve kterém si uživatel zobrazí grafy informací potřebných pro aktuální část výzkumu – například videa s určitou podmínkou (v závislosti na čase, v závislosti na počtu videí) a jejich porovnání.
- Bude zdokonalen systém auditování videí z aplikace TikTok, dle požadavků výzkumu upraven způsob logování.

### **3. Výzkum a návrh technologií:**

- Bude proveden výzkum a bude diskutován výběr vhodných jazyků, knihoven a vývojových prostředí s ohledem na jednoduchost, přenositelnost a výkon.
- Na základě výzkumu a konzultace bude zvolen vhodný přístup k aplikaci, například zdali log graficky zobrazovat real-time, nebo ho oddělovat podle jednotlivých sessions.

### **4. Architektura systému:**

- Aplikace bude navržena tak, aby bylo možné snadno přidávat nové typy výstupních komponent a rozšíření.
- Oddělení vykreslovací logiky, logovacího prostředí a UI bude navrženo s ohledem na další rozšiřitelnost a udržovatelnost.

### **5. Testování a dokumentace:**

- Bude provedeno testování základních funkcí aplikace – správné vykreslování a zobrazování, správnost dat, ověření s reálnými daty.
- Bude vytvořena uživatelská dokumentace (návod k použití) a vývojová dokumentace (architektura, návrhové vzory).

### **6. Prezentace a obhajoba:**

- Bude připravena prezentace projektu, včetně živé ukázky nebo videa, které předvede vytváření a spuštění jednoduché simulace.
- Bude připravena obhajoba projektu včetně představení architektury, výběru technologií a vývoje uživatelského rozhraní.

## **Hodnocení:**

Projekt bude hodnocen na základě následujících kritérií:

- funkčnost a přehlednost výstupního prostředí a reálnost dat,
- srozumitelnost a použitelnost systému pro cílovou skupinu (výzkumníci),
- kvalita návrhu architektury a výběru technologií,
- úroveň dokumentace a prezentace,
- přínos a použitelnost v projektu AI-Auditology.

Stručný časový harmonogram (s daty a konkretizovanými úkoly):

- **Září:** Čerpání informací o výzkumu, tvorba zadání a jeho revize
- **Říjen:** Návrh vhodného datového modelu a jeho konzultace s odborníky
- **Listopad-prosinec:** Návrh a první úpravy logování videí, potřebná úprava aktuální verze aplikace pro zaznamenávání logů, konzultace vhodné datové reprezentace na grafu
- **Leden:** Implementace modulu pro grafické zobrazení, tvorba výstupního prostředí vhodné pro výzkum
- **Únor:** Ladění, optimalizace, zapojení do prvního pilotního výzkumu (bude konzultováno s vedoucím projektu)
- **Březen:** Testování, tvorba dokumentace a příprava prezentace

Prohlašuji, že jsem maturitní projekt vypracoval(a) samostatně, výhradně s použitím uvedené literatury.

V Pardubicích 24. 3. 2026

.....

*(vlastnoruční podpis)*

# Rozdělení práce

Jednotlivé individuální projekty jsou vzájemně úzce propojeny a společně tvoří ucelený celek. Z tohoto důvodu byla teoretická část rozdělena mezi tři autory následovně:

**3.1–3.3** - Problémy sociálních sítí a koncept algoritmů – Michal Tvaroh

**3.4-3-5** - DSA, aktuální prevence a řešení problémů – Erik Macák

**3.6.** - Aktuální audity a jejich metody – Martin Pytlík

Ostatní pasáže byly vypracovány individuálně v rámci specifických zadání jednotlivých autorů.

## **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 algoritnické 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 platform. Praktická část představuje tříložkový 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

algoritnický 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.....	19
2. Použité technologie .....	20
2.1. Knihovny .....	20
2.1.1 UIAutomator2 .....	20
2.1.2 Pydantic.....	20
2.1.3 yt-dlp.....	21
2.1.4 OpenCV .....	21
2.1.5 ShazamAPI .....	21
2.1.6 Pandas .....	21
2.1.7 Numpy.....	21
2.1.8 Matplotlib.....	22
2.1.9 Seaborn .....	22
3. Teoretická část .....	23
3.1. Problémy sociálních sítích .....	23
3.1.1. Dezinformace a manipulativní obsah .....	23
3.1.2. Obsah škodlivý pro duševní zdraví .....	23
3.1.3. Selhání moderace jako systémový problém .....	24
3.2. Teoretický popis fungování algoritmů.....	25
3.3. Psychologické aspekty užívání a rizika algoritmů.....	26
3.4. Prevence a legislativy DSA .....	27
3.4.1. Vymáhání a sankce.....	27
3.4.2. Transparentnost doporučovacích systémů a reklamy.....	28
3.4.3. Právo na odvolání a mimosoudní řešení sporů.....	28
3.4.4. Ochrana nezletilých jako priorisa DSA.....	28
3.4.5. Ověřování věku .....	28
3.4.6. Safety-by-design a doporučovací systémy pro nezletilé .....	28
3.5. Jak sociální sítě plní/neplní legislativní požadavky.....	29
3.5.1. Transparentnost doporučovacích algoritmů v praxi.....	29
3.5.2. Problémy, které sociální sítě aktivně neřeší .....	30
3.6. Auditování algoritmů: metodologie, stav a nedostatky .....	30
3.6.1. Algoritmické auditování jako možné řešení.....	30
3.6.2. Metodologické přístupy k auditu algoritmů .....	31
3.6.3. Dosavadní audity: co bylo prověřeno a zůstává stranou .....	33
3.6.4. Nedostatky: nejednotnost, definice a problém metodologické standardizace.....	34

4. Cíle práce .....	36
4.1. Mobilní agent pro interagování s platformou Instagram reels.....	36
4.2. Prediktivní systém pro vyhodnocování uživatelských interakcí.....	37
4.3. Systém pro vizualizaci datových záznamů .....	38
5. Výzkumná část .....	39
5.1. Mobilní agent.....	39
5.2. Prediktivní systém.....	40
5.3. Vizualizační nástroj .....	41
6. Výsledky .....	42
6.1. Mobilní agent.....	42
6.2. Prediktivní systém.....	44
6.3. Vizualizační nástroj .....	47
6.4. Pilotní audit.....	48
6.5. Limity.....	48
7. Závěr .....	49
8. Seznam použité literatury.....	50
9. Seznam obrázků .....	54
10. Seznam příloh .....	55

# 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. 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.

### 2.1.3. 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.

### 2.1.4. 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.

### 2.1.5. 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.

### 2.1.6. 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.

### 2.1.7. 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.

### 2.1.8. 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 a pandas.

### 2.1.9. 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.

## 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 platform 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

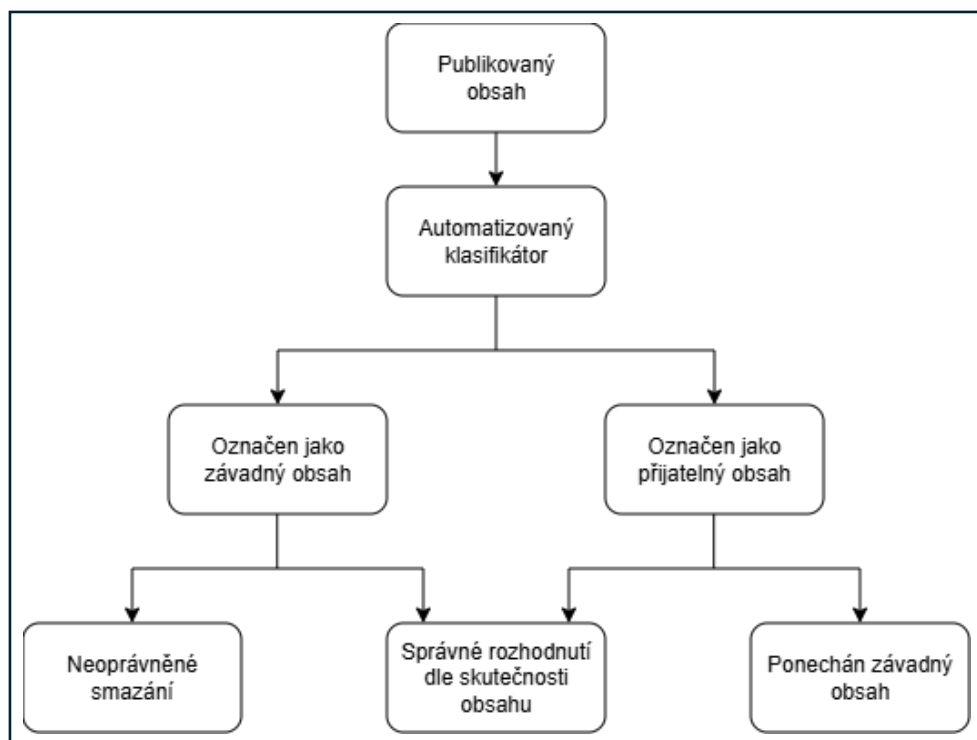
Samostatnou kategorií je obsah, jehož škodlivost není primárně spjata s nepravdivostí, ale s psychologickým účinkem na příjemce. Doporučovací algoritmy mohou děti a adolescenty postupně vystavovat nevhodným či škodlivým materiálům; tento mechanismus je o to závadnější, neboť platformy jsou navrženy primárně 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 sebepoškození 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. (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 platformou ukázala, že hyperpersonalizace 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 nabitě a dělicí 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 platform. 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) – dohlíží na menší poskytovatele ve svém členském státě. (European Commission 2025) Za nedodržení povinností mohou platformám hrozit pokuty až do výše 6 % jejich celosvětového ročního obrátu. (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 platform 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 platform 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 společenskovední 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.

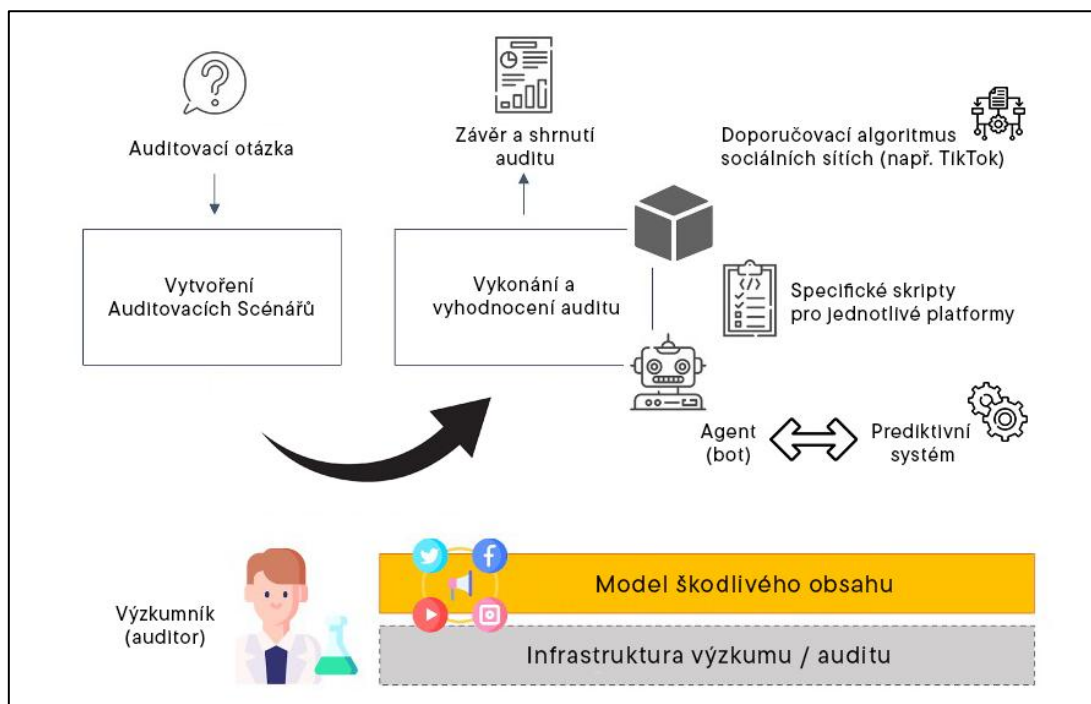
### 3.6.2. Metodologické přístupy k auditu algoritmů

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. (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 audits 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 audits 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 algoritmičké 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 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 doporučovacíh 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 algoritmičtým 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 algoritmičtých 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 platform – 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 platformy 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 systém pro automatizované algoritmičké audity doporučovacích feedů sociálních sítí, který překonává metodologické nedostatky identifikované v dosavadní auditní praxi – zejména absenci standardizace, neschopnost zachytit časovou proměnlivost algoritmů 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, návrh prediktivního systému simulujícího rozhodování syntetických uživatelských profilů a návrh systému pro vizualizaci a interpretaci výsledných auditních dat.

### 4.1. Mobilní agent pro interagování s platformou Instagram reels

Jedním z hlavních praktických cílů této práce je navrhnout a realizovat softwarového agenta, který bude schopen automaticky ovládat mobilní zařízení a provádět interakce v aplikaci Instagram. Tento agent představuje základní stavební prvek celého systému, jelikož zajišťuje samotné vykonávání akcí v prostředí mobilní aplikace. Klíčovým problémem není pouze samotné provedení jednotlivých operací, ale především nalezení spolehlivého a obecně použitelného způsobu, jak simulovat chování reálného uživatele, aby bylo dostatečně přesné, minimalizovalo odchylky od přirozeného uživatelského chování a zároveň bylo odolné vůči změnám uživatelského rozhraní aplikace. Cílem je proto identifikovat vhodné principy pro návrh takového agenta, zejména s ohledem na stabilitu, autenticitu a schopnost reagovat na dynamické prostředí mobilní aplikace.

Druhým praktickým cílem je analyzovat a navrhnout metodologii ovládání mobilního zařízení a interakce s uživatelským rozhraním aplikace Instagram. Na rozdíl od tradičních softwarových systémů zde agent nepracuje s přímým API rozhraním, ale s vizuální vrstvou aplikace, což přináší specifické technické výzvy, jako je rozpoznávání prvků na obrazovce, práce se souřadnicemi nebo interpretace změn v rozhraní. Cílem je proto navrhnout takový přístup, který umožní konzistentní a opakovatelné vykonávání akcí (např. procházení obsahu, interakce příspěvků či správa účtu) nezávisle na konkrétní implementaci aplikace nebo zařízení. Výsledkem by měl být obecný návrhový rámec, který bude možné využít jako základ pro vytvoření funkčního prototypu agenta i pro jeho další rozšiřování.

## 4.2. Prediktivní systém pro vyhodnocování uživatelských interakcí

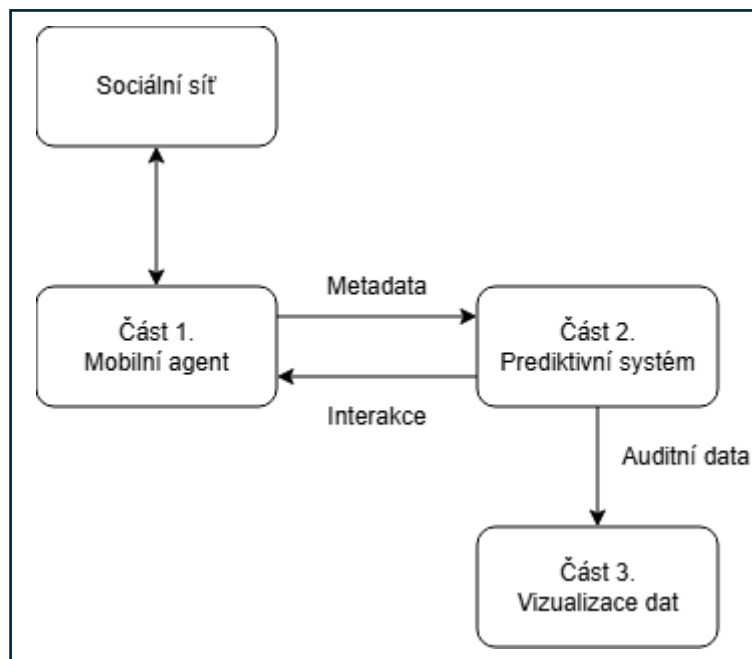
Dalším praktickým cílem této práce je navrhnout a realizovat prediktivní systém, který bude schopen rozhodovat o uživatelských interakcích v prostředí doporučovacích feedů sociálních sítí. Tento systém představuje klíčový nástroj pro algoritmičké audity, jelikož umožňuje testovat chování doporučovacích algoritmů prostřednictvím syntetických uživatelských profilů s přesně definovanými vlastnostmi a preferencemi. Klíčovým problémem není pouze samotné vygenerování jednotlivých akcí, ale především nalezení spolehlivého způsobu, jak v reálném čase analyzovat obsah videa – segmentace videa, vizuální snímky, hudbu na pozadí, metadata o popularitě, textové popisky – a na jejich základě predikovat pravděpodobnou reakci konkrétního typu uživatele. Tato predikce musí být dostatečně přesná, aby věrohodně simulovala skutečné rozhodování, dostatečně rychlá, aby odpověď přišla v řádu sekund, a zároveň dostatečně strukturovaná, aby umožnila srovnání výsledků napříč různými auditovacími scénáři a platformami. Návrh systému reaguje na metodologická selhání identifikovaná v první vlně DSA auditů – především na absenci standardizovaných definic toho, co se vlastně měří a na nesrovnatelnost auditních zpráv mezi platformami. Cílem je proto identifikovat vhodné principy pro návrh takového prediktivního systému, zejména s ohledem na rychlost, přesnost, opakovatelnost a schopnost pracovat transparentně s veřejně dostupnými daty.

Druhým cílem je navrhnout metodologii pro strukturované ukládání a export auditních dat tak, aby výsledky byly použitelné pro následnou interpretaci a srovnání. Na rozdíl od tradičních auditů, kde se data vztahují k relativně statickým procesům, zde systém pracuje s dynamickým obsahem doporučovacích algoritmů, což přináší specifické výzvy – například potřebu zaznamenávat vztah mezi tematickým zařazením obsahu a jeho zacílením na konkrétního uživatele. Současné DSA audity produkují zprávy s nesourodými kategoriemi a metrikami, které nelze vzájemně porovnat. Cílem je proto navrhnout takový datový model a způsob exportu, který umožní konzistentní reprezentaci výsledků nezávisle na konkrétní platformě nebo auditní metodě. Výsledkem by měl být obecný návrhový rámec, který poskytne základ pro vytvoření funkčního systému schopného produkovat auditní data s vědeckou validitou, aniž by vyžadoval přímý přístup k interním datům platformy.

### 4.3. Systém pro vizualizaci datových záznamů

Cílem je identifikovat principy správné prezentace výstupních dat z algoritnického auditu tak, aby byla srozumitelná a použitelná pro různé skupiny příjemců – od výzkumníků přes regulátory až po veřejnost. Způsob zobrazení dat totiž přímo ovlivňuje, jaké závěry z nich lze vyvodit, a proto je třeba stanovit, jaké vlastnosti by správná reprezentace měla splňovat z hlediska přesnosti, přehlednosti a odolnosti vůči dezinterpretaci.

Druhým cílem je zhodnotit metodologické přístupy vhodné pro vizualizaci auditních dat, která jsou – na rozdíl od dat z finančních auditů – dynamická, vícerozměrná a obtížně srovnatelná napříč platformami. Výsledkem by měl být metodologický rámec umožňující konzistentní reprezentaci výsledků nezávisle na konkrétní platformě nebo použité auditní metodě.



Obrázek 3: Schéma celého nástroje

## 5. Výzkumná část

### 5.1. Mobilní agent

Cílem této části bylo navrhnout a implementovat agenta schopného věrohodně simulovat chování reálných uživatelů v prostředí aplikace Instagram. Klíčovým požadavkem bylo zajistit, aby simulované interakce odpovídaly přirozenému uživatelskému chování, jelikož právě míra realismu má zásadní vliv na validitu výsledků algoritmického auditu. Za tímto účelem byla nejprve provedena analýza aplikace, zaměřená na strukturu uživatelského rozhraní a identifikaci možných interakcí. Paralelně proběhlo zkoumání běžných vzorců chování uživatelů při konzumaci krátkých videí. Na základě těchto poznatků byl definován požadavek na implementaci nepravidelných časových rozestupů mezi akcemi, simulaci přirozeného pohybu v prostředí aplikace (např. zakřivené swipování) a zařazení náhodných pauz či doplňkových interakcí. Cílem těchto opatření je minimalizovat odlišitelnost agenta od reálného uživatele.

Dalším cílem bylo propojit agenta s prediktivní částí systému. Agent je proto navržen jako exekuční vrstva, která realizuje rozhodnutí generovaná prediktivním systémem v reálném prostředí aplikace. Prediktivní systém na základě vlastností zobrazovaného obsahu odhaduje pravděpodobnosti jednotlivých uživatelských akcí, přičemž agent tyto výstupy převádí na konkrétní interakce. Součástí návrhu je také záznam provedených interakcí včetně kontextových informací (např. čas, typ akce, vlastnosti videa). Tato data slouží jako vstup pro analytickou část systému a umožňují vyhodnocení chování doporučovacího algoritmu v průběhu jednotlivých session.

V neposlední řadě bylo cílem umožnit simulaci různých typů uživatelů prostřednictvím definovaných uživatelských profilů.

Tyto profily reprezentují odlišné preference, demografické charakteristiky (např. věk či pohlaví) a vzorce chování, a ovlivňují rozhodovací proces agenta, což umožňuje analyzovat dopady personalizace na různé skupiny uživatelů.

## 5.2. Prediktivní systém

Prvním krokem u vývoje prediktivního systému bylo nutné provést podrobnou analýzu několika oblastí týkajících se fungování doporučovacích algoritmů a uživatelského chování. Tato analýza se zaměřovala především na identifikaci klíčových faktorů ovlivňujících uživatelské rozhodování při konzumaci krátkých videí. Součástí bylo zkoumání již existujících přístupů a nástrojů k algoritmickému auditu – zejména jejich interního fungování, hlavních limitů a návrhů možných zlepšení. Dalším krokem byla analýza běžných vzorců uživatelského chování v prostředí „For You“ feedů – konkrétně podmínek, za kterých uživatelé typicky vykonávají určitý druh interakce. Z této analýzy vyplynula potřeba definovat strukturované auditovací scénáře prostřednictvím uživatelských profilů obsahujících preference jako témata zájmu, preferované emoce, jazyky či oblíbené autory. Paralelně byla provedena analýza nejdůležitějších vlastností video obsahu ovlivňujících interakci – vyvolávané emoce, technická kvalita, přítomnost známé hudby, jazyk mluveného slova nebo téma videa. Přesnost těchto definic má přímý vliv na věrohodnost simulovaného chování a tím i na validitu výsledků auditu.

Klíčovým metodologickým rozhodnutím bylo zavedení segmentace videa, kdy místo analýzy celého obsahu jsou analyzovány pouze vybrané části v závislosti na jeho délce. Tento přístup vychází z poznatku, že uživatelé se typicky rozhodují v prvních několika sekundách sledování, a proto není nutné (ani časově efektivní) analyzovat celé video. Pro videa různých délek byly definovány specifické strategie. Tento přístup vyžadoval analýzu toho, jaké segmenty jsou z hlediska rozhodování nejrelevantnější – typicky začátek videa, případně přechody mezi sekvencemi. Klíčovým požadavkem byla rychlost analýzy v řádu sekund. Toho bylo dosaženo optimalizací celého procesu, od snížení frekvence extrakce snímků až po volbu efektivnější architektury AI modelů pro detekci specifických vlastností videa.

### 5.3. Vizualizační nástroj

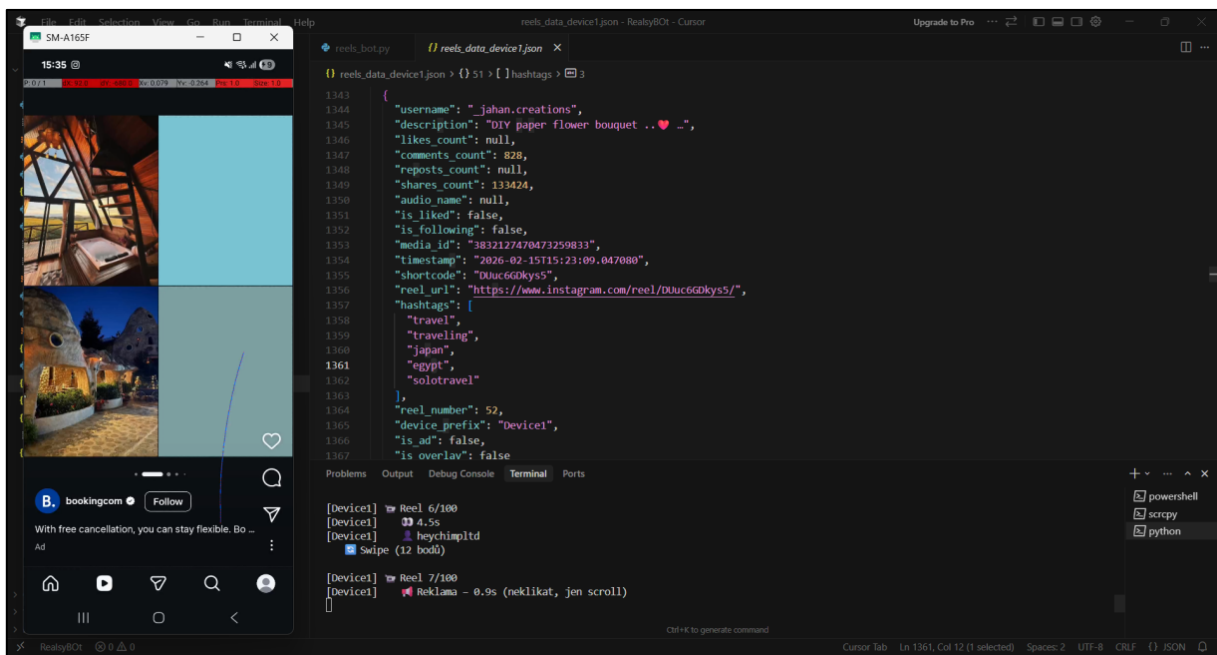
Výchozím krokem při návrhu vizualizačního nástroje bylo stanovení požadavků, které musí datová reprezentace výsledků algoritmičského auditu splňovat. Z povahy auditních dat vyplynuly tři klíčové vlastnosti. Za prvé **konzistence** – všechny vizualizace musejí sdílet jednotný vizuální jazyk, aby bylo možné výsledky porovnávat napříč tématy, postoji a časovými úseky bez nutnosti přizpůsobovat interpretaci každému grafu zvlášť. Za druhé **úplnost pohledu** – každá vizualizace musí zobrazovat jak relativní podíly, tak absolutní hodnoty, protože nevyvážené skupiny mohou při zobrazení pouze procentuálních hodnot vést k dezinterpretaci. Za třetí **srozumitelnost pro různé skupiny příjemců** – výstupy musejí být čitelné jak pro technicky orientované výzkumníky, tak pro regulátory bez hlubokého datového zázemí, což klade nároky na popis os, anotace hodnot a volbu typu grafu.

Na základě stanovených požadavků byly pro jednotlivé analytické úlohy zvoleny konkrétní typy vizualizací. Pro zobrazení celkové distribuce dat byla kombinována relativní (donut chart) a absolutní (horizontální bar chart) reprezentace vedle sebe – tento přístup zamezuje ztrátě kontextu při nevyvážených skupinách. Pro sledování vývoje složení feedu v průběhu session byl zvolen stacked area chart s klouzavým průměrem (sliding window), který vyhlazuje krátkodobý šum a odhaluje trendy personalizace. Pro srovnání více metrik napříč tématy byl použit radar chart, který umožňuje vizuální srovnání vícerozměrných profilů bez nutnosti číst sadu oddělených grafů. Distribuce predicted témat v čase byla binována pomocí pd.qcut, který zajišťuje přibližně stejný počet záznamů v každém binu – na rozdíl od pd.cut, který vytváří stejně široké intervaly bez ohledu na rozložení dat. Klíčovým metodologickým rozhodnutím bylo zavedení tzv. Bubble Indexu – ukazatele, který měří podíl videí potvrzujících postoj uživatele výhradně ze skupiny videí s matchujícím tématem. Tím je efekt stance filtrování izolován od obecné tematické personalizace, což zvyšuje interpretační přesnost výsledku.

# 6. Výsledky

## 6.1. Mobilní agent

Výsledkem implementace agenta je Python skript využívající knihovnu UiAutomator2, která umožňuje ovládat aplikace na fyzických Android zařízeních a simulovat reálné uživatelské interakce. Script provádí typické uživatelské akce na platformě Instagram Reels, včetně procházení „Reels feedu“, přehrávání videí, lajkování a dalších interakcí, přičemž zachovává přirozenou variabilitu chování pomocí nepravidelných časových rozestupů, zakřiveného swipování a náhodných pauz. Každé video je identifikováno svým video ID, načítaným z lokální databáze SQLite, nebo přímo z dat dostupných v zařízení během běhu aplikace na rootnutém telefonu, a spolu s dalšími relevantními údaji je odesíláno do prediktoru, který je vyhodnotí a vrátí doporučenou akci; agent následně tuto akci provede ve feedu. Všechna nasbíraná data mohou být následně využita pro tvorbu vizualizací a analýzu chování systému.



Obrázek 4: Ukázka promítnutého zařízení, sběr metadat z Instagram Reels

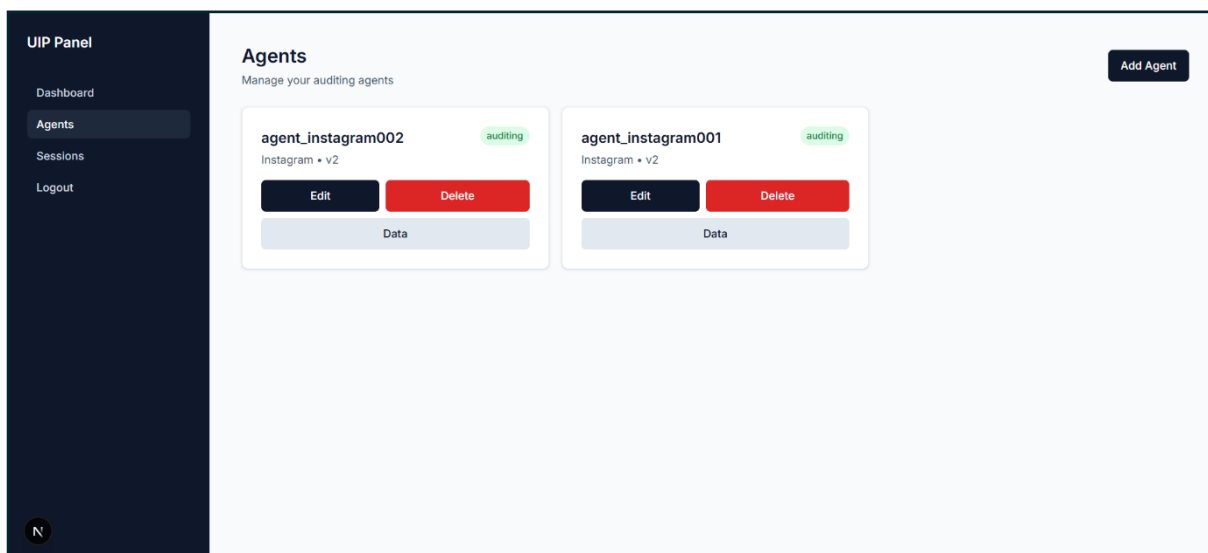
## Ověření a diskuse

Agent byl ověřen na reálných zařízeních s rootnutým Androidem pomocí testovací sady účtů a předem definovaných uživatelských profilů. Testy zahrnovaly simulaci typických interakcí ve „For You“ feedu, přehrávání videí, lajkování a dalších akcí, přičemž byly monitorovány časové rozestupy, pohyb ve feedu a náhodné pauzy, aby byla ověřena věrohodnost chování. Výsledky ukázaly, že agent dokáže provádět interakce v souladu s prediktorem a generuje data odpovídající realistickému uživatelskému chování, což potvrzuje efektivitu návrhu při simulaci variability akcí i různých uživatelských profilů.

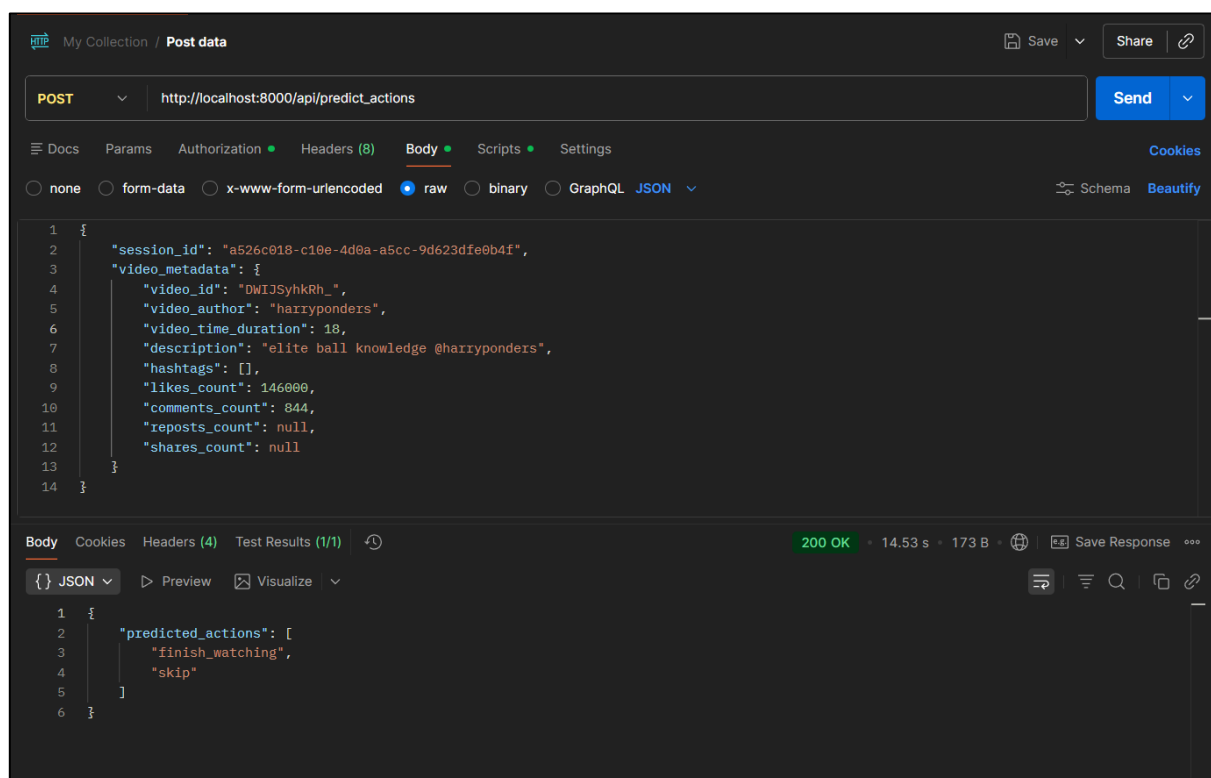
Limitem ověřování je, že testy probíhaly pouze na omezeném počtu účtů a videí, takže úplná generalizace na širší populaci uživatelů není bez dalšího potvrzení možná. Dalším omezením je závislost na stabilitě rootnutého zařízení a přesnosti přístupu k lokální databázi, která může ovlivnit spolehlivost získaných video ID a souvisejících dat.

## 6.2. Prediktivní systém

Výsledkem implementace prediktivního systému je webová aplikace s backendem v Python frameworku FastAPI a frontendem v Next.js, sloužící pro administraci agentů a auditovacích session, přičemž primární funkcionalitou je veřejný API endpoint schopný v reálném čase vypočítat a vrátit optimální uživatelskou interakci na základě poskytnutých metadat, analýzy multimediálního obsahu a auditovacích scénářů konkrétního agenta. Aplikace implementuje autentizaci a autorizaci prostřednictvím JWT tokenů, kdy při prvním spuštění je uživatel vyzván k vytvoření silného hesla, které následně slouží pro přístup do systému. Jádrem administračního rozhraní je správa agentů (CRUD operace), která uživateli umožňuje přizpůsobit jednotlivé agenty konkrétním auditovacím scénářům – vybrat cílovou platformu, poskytnout strukturovaný uživatelský profil obsahující preference (témata, emoce, jazyky, oblíbené autory), zvolit výpočetní verzi prediktoru a exportovat nasbíraná auditní data vztahující se ke konkrétnímu agentovi. Dalším klíčovým modulem je správa auditovacích session, kdy uživatel vybere agenta, se kterým chce provádět audit, a systém pro něj vytvoří unikátní identifikátor session (`session_id`), který slouží jako autentizační token pro volání prediktivního API. Tento identifikátor je platný pouze po dobu trvání aktivní session a po jejím ukončení je zneplatněn, čímž je zajištěno, že k provolání prediktoru nemůže přistupovat neautorizovaná strana. Samotný výpočet interakce probíhá na základě poskytnutých metadat o videu (autor, délka, engagement metriky, popisek, hashtagy), analýzy multimediálního obsahu (extrakce vizuálních snímků, detekce hudby pomocí Shazam API, AI-poháněné rozpoznání emocí, témat a technické kvality) a rozhodovacího stromu reflektujícího zjištěné vzorce uživatelského chování v prostředí „for you“ feedů. Výstupem je sekvence akcí, která věrohodně simuluje rozhodování reálného uživatele s daným profilem.



Obrázek 5: Screenshot webového rozhraní prediktivního systému, slouží pro revizi aktuálně běžících auditů a pro export dat



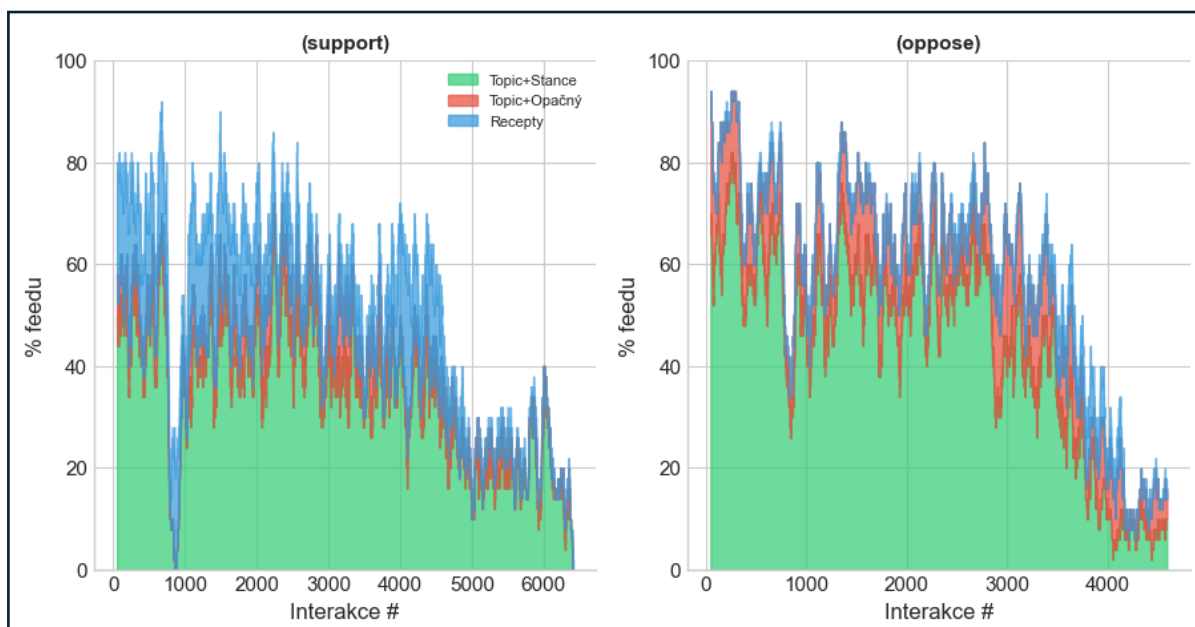
Obrázek 6: Demonstrace klientského dotazu pro výpočet uživatelské interakce

## **Ověření a diskuse**

Prediktivní systém byl ověřen pomocí testovací sady syntetických uživatelských profilů s různými preferencemi a auditovacími scénáři. Testy zahrnovaly primárně konzistenci predikovaných interakcí s definovanými uživatelskými profily. Dále byla ověřována rychlost odezvy systému při zpracování videí různých délek, přičemž průměrný čas analýzy se pohyboval v rozmezí 5-30 sekund napříč různými délkami videí, což odpovídá realistickému uživatelskému chování při rozhodování. Výsledky ukázaly, že systém dokáže spolehlivě, rychle a opakovatelně vracet interakce v souladu s poskytnutými auditovacími scénáři a produkovat strukturovaná data vhodná pro následnou analýzu a vizualizaci, což potvrzuje efektivitu návrhu při simulaci autentického uživatelského chování bez nutnosti přístupu k interním datům platformy.

### 6.3. Vizualizační nástroj

Výsledkem praktické části je Jupyter notebook implementovaný v jazyce Python, který přijímá na vstupu CSV soubor se záznamy z algoritmického auditu a produkuje sadu čtrnácti vizualizací uložených do výstupní složky spolu se souhrnnou tabulkou ve formátu CSV. Nástroj je členěn do tematických sekcí odpovídajících analytickým vrstvám auditu: globální přehled datasetu, analýza algoritmické bubliny, analýza interakcí, přesnost algoritmu, srovnání témat demografická analýza. Vizuální konzistence je zajištěna centrálně definovanou barevnou paletou a globálním nastavením parametrů knihovny Matplotlib, takže změna stylu nevyžaduje úpravy v jednotlivých buňkách. Každá buňka řeší jednu analytickou otázku a je opatřena markdown popisem, což usnadňuje orientaci i bez znalosti kódu. Výstupy jsou navrženy tak, aby byly použitelné přímo v prezentacích a zprávách bez další úpravy.



Obrázek 7: Vývoj složení feedu v průběhu sessions na platformě (Sliding window = 50) v rámci tématiky US Politics. Jedná se o data s testovacího datasetu poskytnutá institutem KInIT pro zjednodušení práce a příprav vyhodnocovacího systému.

## 6.4. Pilotní audit

Nástroj je ověřován 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. Z důvodu oprav chyb a integrace celého systému aktuální audity zatím probíhají.

Mezi aktuálním vzorkem dat zatím nebyl na platformě Instagram reels nalezen žádný vzorec promování rizikového obsahu u uživatelů, kteří nevhodný obsah vyhledali explicitně. Instagram naopak doporučuje a upozorňuje na vyhledání takového obsahu přímo v aplikaci, ať už cenzurou příspěvků, nebo upozorněním přímo v aplikaci.

## 6.5. Limity

Vyvinutý systém přináší funkční základ pro automatizované algoritnické 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. 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 postuje ve vstupních datech – chyby v předřazené klasifikaci se přímo promítají do výsledků auditu, aniž by je nástroj sám zachytil. 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, neschopnost zachytit časovou proměnlivost algoritmů a nesrovnatelnost výsledků mezi platformami. Navržený tříložkový systém – mobilní agent, prediktivní systém a vizualizační systém – mobilní cíl naplňuje: umožňuje opakovatelné a strukturované audity bez nutnosti přístupu k interním datům platform. 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 platform, implementaci statistického testování hypotéz ve vizualizačním modulu a dlouhodobé testování odolnosti agenta vůči změnám uživatelského rozhraní.

## 8. Seznam použité literatury

ADA LOVELACE INSTITUTE, 2021. *Technical methods for regulatory inspection of algorithmic systems*. Online. Nevedeno: 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](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-long-awaited-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](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](https://doi.org/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](https://doi.org/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](https://doi.org/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. Nevedeno: 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. Nevedeno: 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.stepto.com/en/news-publications/stepteptoe-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

## 9. Seznam obrázků

OBRÁZEK 1: DIAGRAM ROZHODOVÁNÍ O OBSAHU .....	24
OBRÁZEK 2: METODOLOGIE SOCK-PUPPET AUDITU .....	32
OBRÁZEK 3: SCHÉMA CELÉHO NÁSTROJE .....	38
OBRÁZEK 4: UKÁZKA PROMÍTNUTÉHO ZAŘÍZENÍ, SBĚR METADAT Z INSTAGRAM REELS.....	42
OBRÁZEK 5: SCREENSHOT WEBOVÉHO ROZHRAŇÍ PREDIKTIVNÍHO SYSTÉMU, SLOUŽÍ PRO REVIZI AKTUÁLNĚ BĚŽÍCÍCH AUDITŮ A PRO EXPORT DAT .....	45
OBRÁZEK 6: DEMONSTRACE KLIENTSKÉHO DOTAZU PRO VÝPOČET UŽIVATELSKÉ INTERAKCE .....	45
OBRÁZEK 7: VÝVOJ SLOŽENÍ FEEDU V PRŮBĚHU SESSIONS NA PLATFORMĚ (SLIDING WINDOW = 50) V RÁMCI TÉMATIKY US POLITICS. JEDNÁ SE O DATA S TESTOVACÍHO DATASETU POSKYTNUTÁ INSTITUTEM KINIT PRO ZJEDNODUŠENÍ PRÁCE A PŘÍPRAV VYHODNOCOVACÍHO SYSTÉMU. ....	47

## 10. Seznam příloh

PŘÍLOHA: DOKUMENTACE API .....	I
--------------------------------	---

# Příloha: Dokumentace API

Následující přehled shrnuje hlavní endpointy backendového API prediktivního systému. U každé položky je uvedena metoda, cesta, základní vstup a očekávaný výstup. Přehled je určen jako provozní a integrační dokumentace pro práci s aplikací.

Metoda	Endpoint	Vstup	Návratová hodnota
POST	<i>/api/login</i>	password	zpráva, informace o nutnosti změny hesla
POST	<i>/api/change-password</i>	new_password, confirm_password	potvrzení o změny hesla
POST	<i>/api/logout</i>	-	potvrzení odhlášení + smazání cookies
GET	<i>/api/verify</i>	auth_cookie	stav přihlášení
GET	<i>/api/platforms</i>	-	seznam podporovaných platform
GET	<i>/api/predictor-versions</i>	-	seznam výpočetních verzí prediktivního systému
POST	<i>/api/agents</i>	AgentCreateRequest	vytvořený agent
GET	<i>/api/agents</i>	state (nepovinné)	seznam agentů
GET	<i>/api/agents/{agent_id}</i>	agent_id	detail konkrétního agenta
PUT	<i>/api/agents/{agent_id}</i>	AgentUpdateRequest	aktualizovaný agent
DELETE	<i>/api/agents/{agent_id}</i>	agent_id	potvrzení o smazání agenta
POST	<i>/api/sessions</i>	agent_id	vytvořená session
GET	<i>/api/sessions/running</i>	-	seznam běžících sessions
POST	<i>/api/sessions/{session_id}/stop</i>	session_id	potvrzení zastavení sessions
POST	<i>/api/predict_actions</i>	PredictActionsRequest	seznam predikovaných akcí
GET	<i>/api/agents/{agent_id}/sessions</i>	agent_id	přehled sessions agenta
GET	<i>/api/agents/{agent_id}/export</i>	agent_id	export všech dat agenta nasbírané v průběhu sessions