

MUNI
FI

Filtrování informací a cenzura

Mgr. Tomáš Foltýnek, Ph.D.

foltynec@fi.muni.cz

Osnova dnešní přednášky

- Opakování: Mediální manipulace a dezinformace na internetu
- Téma: Filtrování informací a cenzura
 - Nová Velká čínská zeď
 - Právo být zapomenut + diskuse
 - Blokování dezinformačních webů + diskuse
 - Filtrování v univerzitní síti
- Dilema game
 - Filtrování v univerzitní síti
 - Citlivá data

OPAKOVÁNÍ: Mediální manipulace a dezinformace na internetu



Opakování: Terminologie

- Dezinformace
 - **Záměrně falešná informace** s cílem ovlivnit názory lidí
 - Vytvořena tak, aby **vyvolávala zdání pravdivosti**
 - Vzniká manipulací existující (pravdivé) zprávy nebo vytvořením zcela nové zprávy
- Misinformace
 - Neúmyslně chybná zpráva
 - Satira
 - Tj. bez úmyslu ovlivnit něčí názor

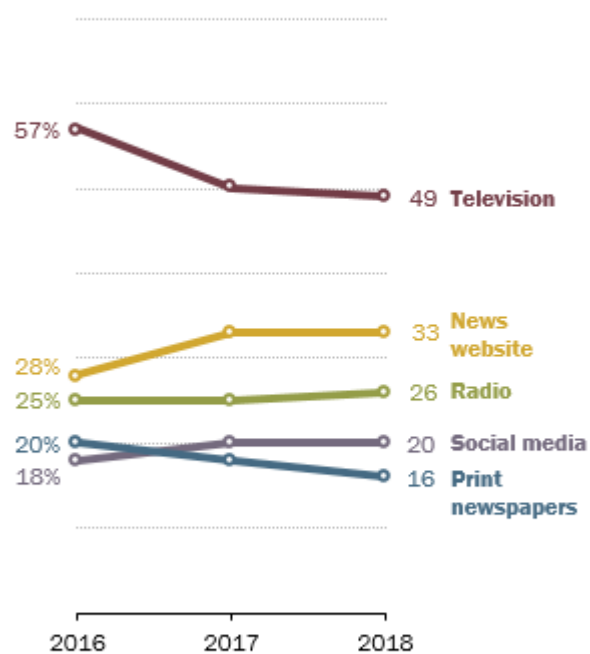
Opakování: Psychologická podstata

- Proč lidé věří dezinformacím?
 - **Naivní realismus**: Lidé věří, že jejich vnímání reality je jediné správné
 - Potvrzovací zkreslení (**confirmation bias**): Lidé preferují informace, které potvrzují jejich vidění světa
- Je velmi obtížné důvěru narušit
 - Oprava mylných údajů (pokud pochází “z opačného tábora”) může dokonce ještě více pokřivit vnímání reality
- Lidé se snaží maximalizovat osobní užitek
 - Včetně společenského přijetí a ocenění
 - Co s podivným příspěvkem od kamaráda?
 - Lajkovat a sdílet, nebo ověřovat a vyvracet?
 - Která volba je “společensky bezpečnější”?

Opakování: Zdroje informací

More Americans get news often from social media than print newspapers

% of U.S. adults who get news *often* on each platform



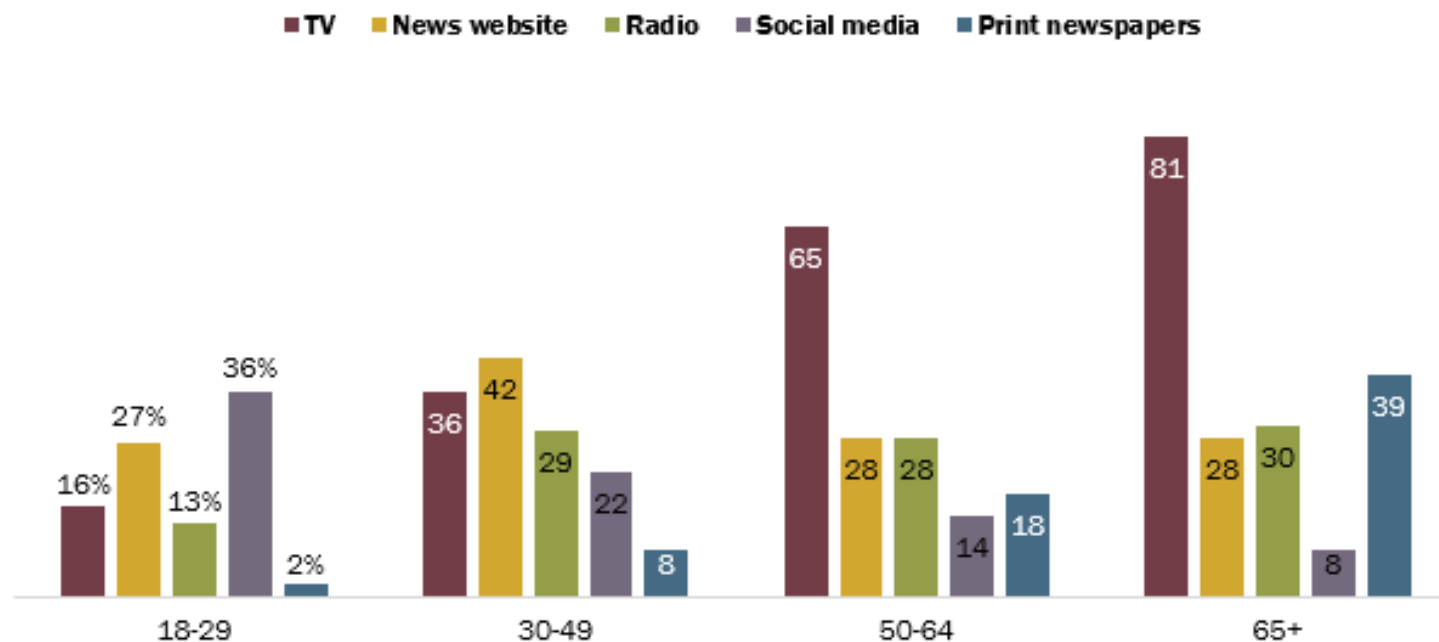
Note: The difference between social media and print newspapers in 2017 was not statistically significant.

Source: Survey conducted July 30-Aug. 12, 2018.

PEW RESEARCH CENTER

Television dominates as a news source for older Americans

% of each age group who *often* get news on each platform



Source: Survey of U.S. adults conducted July 30-Aug. 12, 2018.

PEW RESEARCH CENTER

Zdroj: <https://www.pewresearch.org/fact-tank/2018/12/10/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/>

Opakování: Zpravodajské servery

- Zamýšlený cíl: Co nejlépe informovat o tom, co se děje
 - Přiměřeně, zodpovědně
- Realita: Optimalizace na analytiku
 - Počet zobrazení / kliků / lajků
- Většina čtenářů kliká na negativní a senzační zprávy
 - Poskytování takovýchto zpráv uspokojuje poptávku
- Vyhrávají nejsenzačnější titulky
 - Útočící na city a pudy čtenářů
- Vytváří podhoubí pro extrémní a manipulativní informace

Opakování: Maximalizace užitku

- Autor zprávy
 - Krátkodobý užitek: Dosah na co nejvíce čtenářů
 - Dlouhodobý užitek: Autenticita zpráv → Reputace
- Čtenář zprávy:
 - Informační užitek: Dozví se pravdivou a nepokřivenou informací
 - Psychologický užitek: Dozví se informací, která zapadá do jeho názorů a potřeb
- Dezinformace se šíří pokud
 - Autoři upřednostňují krátkodobý užitek
 - Čtenáři upřednostňují psychologický užitek

Opakování: Dezinformace na soc. médiích

- Sociální boti (social bots)
 - Programy, které automaticky publikují obsah a interagují s uživateli
 - Odhadu: V den voleb Trump vs. Clinton tweetovalo cca 19 milionů botů
- Trollové
 - Účastníci online diskusí
 - Snaží se záměrně vyvolávat hádky, provokovat, urážet skupiny či jednotlivce, šířit propagandu nebo falešné informace
- Sociální kyborgové
 - Kombinace lidského vstupu a automatických programů na zvýšení dosahu
- Sociální bubliny → Šíření informací, kterým bublina věří → Posilování důvěry → Větší uzavřenost bubliny

Opakování: Šíření falešných zpráv

- [Studie MIT](#) na datech z Twitteru z let 2013-2017
- Publikovaná 2018 v Science
- Falešné zprávy se šíří 3x rychleji
- Nešíří je boti, ale skuteční lidé
 - Někteří záměrně, jiní nevědomky
- Falešné zprávy → Překvapující, nové → Chci se podělit

Opakování: “Going Viral”

- [Výzkum](#) na Stanfordu 2019
- Analogie šíření falešných zpráv a šíření viru
 - Nemocný = věří zprávě
 - Nakažlivý = šíří zprávu dál
- Snaha nalézt nejzranitelnější jedince
 - Podobně jako u viru
 - Čím více se lidé setkávají s falešnými zprávami (navíc z důvěryhodných zdrojů), tím snáze ”onemocní”
- Stejně jako u virů fungují “superšířitelé”
 - Málo lidí s velkým dosahem na sociálních médiích

Opakování: Očkování proti dezinformacím

- Zpravodajské servery
 - Sebereflexe
 - Etika novinářské práce
- Uživatelé
 - Kritické myšlení
 - Ověřování faktů
- Provozovatelé sociálních médií
 - Identifikovat falešné zprávy a “dát je do karantény”
 - To ovšem stojí peníze hned 2x
 - Vývoj a provoz algoritmů na detekci
 - Lidé přijdou o “zajímavý” obsah a budou trávit méně času u obrazovky

Opakování: Autom. detekce dezinformací

- Velmi obtížný úkol
 - Zprávy jsou záměrně napsané tak, aby jim lidé věřili
 - Zprávy mají různý styl a týkají se různých témat
 - Zprávy obvykle citují důvěryhodné zdroje
 - Ale překrucují je nebo zasazují do nepatřičného kontextu
- Modely založené na obsahu
 - Detekce pravdivosti pomocí ověřování tvrzení v článku
 - Manuálně (experti), crowdsourcing, automaticky
- Modely založené na stylu
 - Snaha přesvědčit, manipulovat → Detekce zavádějících titulků
- Modely založené na vyhodnocení postojů uživatelů
 - Vyhodnocení, zda daný uživatel s článkem souhlasí či nikoliv
- Modely založené na vyhodnocení šíření
 - Vyhodnocení důvěryhodnosti pomocí provázanosti mezi příspěvky

Dotazy?

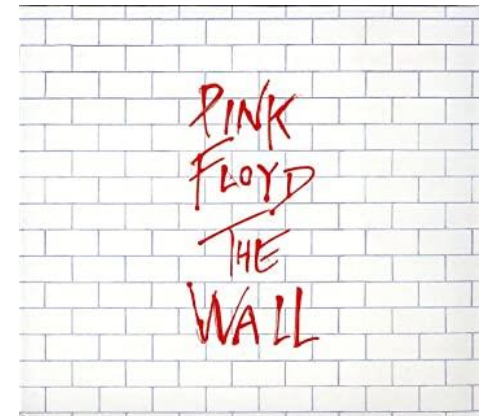
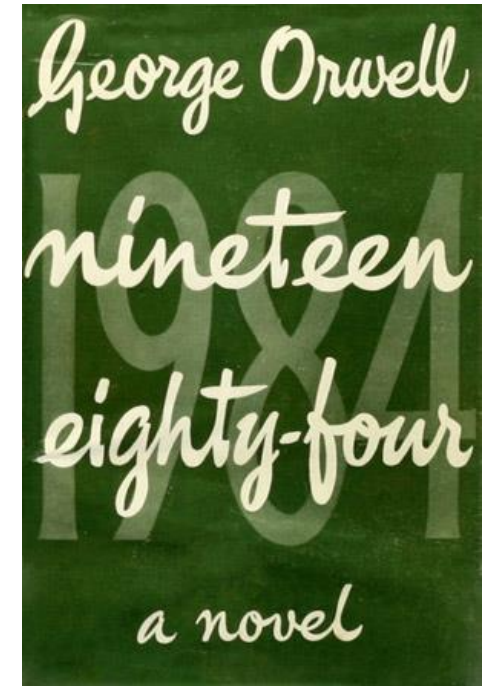
- K předchozí přednášce?
- K analýze využitelnosti Chat GPT ve zvoleném předmětu?

Filtrování informací a cenzura



Nová Velká čínská zed'

- Chrání zemi před "hrozbami" jako např.
 - občanská společnost, svoboda slova
 - falešné představy a myšlenky
- Blokování "závadného" obsahu
 - Včetně celých domén (bbc.co.uk, seznam.cz,...)
 - Filtrování výsledků hledání v Googlu
 - Uvěznění a zmizení autorů nepohodlného obsahu
- Vzniká státem kontrolovaná informační bublina
 - Mnohými (většinou společností?) vnímána pozitivně
 - Pomáhá udržovat současné mocenské zřízení



EU: Právo být zapomenut

- Právo na odstranění odkazů na určité stránky **z internetových vyhledávačů** a dalších adresářů
 - Nikoliv na odstranění obsahu jako takového
- Zohlednění vývoje osobnosti člověka
 - chyby z minulosti, které byly potrestány / napraveny / odpuštěny
 - obvinění, která byla vyvrácena
 - aktivity / události, které již nejsou relevantní
 - mohou vést k opakované stigmatizaci a znevýhodňování
- Hodnoty, které jsou v konfliktu
 - svoboda slova, právo na informace
 - právo na soukromí, ochrana osobnosti
- Google zveřejňuje [aktuální data](#) o počtech požadavků na odstranění odkazů
 - Úspěšná je cca polovina požadavků

Blokování dezinformačních webů

- 25. února 2022 (den po vypuknutí války na Ukrajině) zablokoval CZ.NIC 8 domén dezinformačních webů
 - Aeronet.cz, Protiproud.cz, Ceskobezcenzury.cz, Prvnizpravy.cz, Czechfreepress.cz, Voxpopuliblog.cz, Exanpro.cz a Skrytapravda.cz
- Bylo toto konkrétní rozhodnutí správné?
- Za jakých okolností je blokování konkrétních domén správné?
 - Aneb kde končí svoboda slova?
 - Kdo by o tom měl rozhodovat?

Blokování dezinformačních webů

- ČR připravuje zákon na boj s dezinformacemi (SR už má)
- Rekonstrukce státu spolu s dalšími organizacemi vydali [Doporučení pro vypínání serverů](#)
 - Kritérium konkrétní, bezprostřední a vážné újmy
 - Jasná definice, kdy lze k vypnutí přistoupit
 - Využití širší sady nástrojů
 - Výzva provozovateli, stažení jednotlivého článku,...
 - Soulad s lidskoprávní judikaturou
 - Ochrana demokracie před hrozbami je dvojsečná
 - Řádný proces včetně možnosti odvolání
 - Lhůty, zveřejněné zdůvodnění
 - Pravomoc
 - Mezioborový úřad, soud
 - Posilování role občanské společnosti
 - Vyvracení dezinformací, šíření pravdy

Filtrování v univerzitní síti?

- Problém: Contract cheating
 - Studenti si najmou někoho, kdo za ně vypracuje úkol
 - Typicky seminární či závěrečná práce
 - Je to forma plagiátorství
 - Studenti odevzdají cizí dílo, které vydávají za své vlastní
 - Nejistitelné současnými antiplagiátorskými systémy
 - Problém zejména v anglosaském vzdělávacím systému
 - Znamky ve velké míře závisejí na samostatných úkolech
 - 3-8 % student se přiznává, že se někdy během studia contract cheatingu dopustilo
 - 2/3 to dělají opakovaně → Pro 2-5 % studentů je to standardní modus operandi
- Mnoho firem inzeruje své služby jako „pomoc s psaním práce“
 - Studenti se nechávají nalákat na zdánlivě legální službu

Filtrování v univerzitní síti?

- Některé univerzity blokují přístup na tyto stránky z vnitřní sítě
 - Doporučení např. britské QAA
- Jiné jen zobrazí varování, ale uživatele pustí dále
- Obdobný problém: Chat GPT
 - Některé školy blokují přístup
- Je filtrování správné?
 - Jaká jsou / mají být kritéria?
 - Kdo má rozhodovat o zařazení na seznam blokováných stránek?



MASSEY UNIVERSITY
UNIVERSITY OF NEW ZEALAND

Web Filtering

Powered by ITS Security

Web Page Blocked!

The website you have requested has been deemed as inappropriate. This activity breaches Massey University's Internet access policies. The website has been blocked due to clear, historical breaches of policy.

Site categorisation is not 100% accurate, so if you feel you have received this message in error, please contact the ITS Service Desk via the following [online help request form](#), or by phone on extension 82111.

Internet usage is routinely monitored and logged.

URL: proxyturbo.com/
Category: Proxy Avoidance
Client IP: 130.123.32.165
Server IP: 107.167.86.174
User name:
Group name:

If you have a legitimate study, research or work related reason to access this site, or if you have questions regarding any of the information in this notice, please contact the ITS Service Desk via the following [online help request form](#), or by phone on extension 82111.

Note that permission for this access will have to be authorised by your Head of Department or your manager. Massey University's Internet usage policies for staff and students can be found [here](#).

Dilemma Game



Dilema: Filtrování v univerzitní síti

Mnohé webové stránky představují hrozbu pro návštěvníky. Jedná se o bezpečnostní hrozby (phishing, malware), nelegální obsah, dezinformace, etická rizika, atd. Jak by měla univerzita přistupovat k filtrování těchto hrozeb?

- A. Nijak. Svoboda slova je nade vše, na univerzitu žádné filtrování nepatří.
- B. Filtrovat pouze prokazatelné bezpečnostní hrozby, vše ostatní by mělo být povoleno.
- C. Svoboda slova má své hranice. Univerzita může v závažných a odůvodněných případech výjimečně přistoupit k blokování vybraných stránek či domén.
- D. Univerzita je vzdělávací instituce. Je v pořádku blokovat nevhodné stránky, které uživatelé nepotřebují ke svojí práci nebo studiu.

Dilema: Citlivé výsledky

Brilantní studentka chce psát svoji diplomovou práci o tom, zda a do jaké míry jistá konkrétní sociální síť špehuje své uživatele. Společnost, která stojí za danou sítí, je však sponzorem mnoha projektů na naší fakultě, její zástupci pravidelně přednášejí v různých předmětech, a navíc je zaměstnavatelem mnoha našich studentů.

Pokud dané téma přijmu, očekávám, že tento výzkum nebude vedením fakulty či univerzity příliš vítán a já jako vedoucí budu označen za potížistu, což mě může v budoucnosti poškodit v akademickém postupu. Co mám dělat?

- A. Studentce navrhnou jiné, méně kontroverzní témata pro její práci.
- B. Studentku důrazně upozorním na potenciální negativní důsledky její práce, ale povzbudím ji, aby si své téma obhájila a pracovala na něm, protože je to světově závažné téma a ona má vhodné předpoklady, aby to zvládla.
- C. Zeptám se děkana a řídím se jeho radami o tom, jak dále postupovat v dané situaci.
- D. Studentce řeknu, že souhlasím s daným tématem a budu jejím školitelem, upozorním ji však, že budeme muset být velice opatrní s formulací negativních závěrů, které by mohly poškodit danou společnost a tím tranzitivně i naši univerzitu.

Příští týden

- Téma: Zkreslení v algoritmech
 - Souvislost s etikou umělé inteligence
 - Zkreslení v datech, zkreslení v procesu
 - Důsledky zkreslení – diskriminace
 - Prevence zkreslení
- Úkoly
 - Annette Zimmermann & Bendert Zevenbergen: [AI Ethics: Seven traps](#)
 - Podívat se na film [The Social Dilemma](#)