

M U N I
F I

Mediální manipulace a dezinformace na internetu

Mgr. Tomáš Foltýnek, Ph.D.

foltynek@fi.muni.cz

Osnova dnešní přednášky

- Opakování: Sociální média
- Mediální manipulace a dezinformace na internetu
 - Terminologie, nejčastější zdroje informací
 - Ekonomika zpravodajských serverů
 - Šíření dezinformací na sociálních sítích
 - Detekce dezinformací
- Dilema game:
 - Esej
 - Psát pro své publikum

Opakování: Sociální média – pojmy

- **Sociální síť** = skupina lidí, kteří spolu udržují komunikaci různými prostředky
- **Sociální médium** = nástroj umožňující sdílení a výměnu informací mezi lidmi
- **Komnata ozvěn** (echo chamber) = Skupina osob se vzájemnou komunikací utvrzuje ve svých názorech
- **Sociální bublina** = Imaginární bariéra mezi skupinami osob zamezující vzájemné výměně informací
- **Polarizace společnosti** = Míra nesouhlasu s jinými stranami (zatímco vlastní strana je názorově homogenní)

Opakování: Soc. bubliny a polarizace spol.

- Argumenty pro
 - Polarizace je způsobena nedostatečným pochopením rozdílných názorů
 - Echo chambers utvrzují své členy v jejich názorech
 - Opačný názor je považován za nabezpečný pro společnost
 - Sociální bubliny filtrují rozdílné názory
- Argumenty proti
 - Lisé si vždy vybírali (dávno před internetem a soc. médii)
 - Samostatné vs. algoritmické filtrování
 - Uživatelé sociálních médií mají pestřejší zdroje zpráv
 - Čtení zpráv názorové protistrany posiluje polarizaci

Opakování: Kapitalismus dohledu

- Shoshana Zuboff: Surveillance capitalism
- Soukromé společnosti vytvářejí zisk pomocí informací o svých uživateliích
 - Poskytovaná služba není smyslem jejich existence
 - Je prostředkem k získání co nejvíce informací
 - Informace jsou prostředkem k vytváření zisku
- Pasivní využívání našeho nitra → Jeho aktivní ovlivňování
 - Nálada uživatelů, na co kliknout, co si koupit, koho volit...
- Za účelem zisku dochází k “dálkovému ovládnání lidí”
 - Ohrožení lidské svobody a demokracie
- Řešení: Regulace, tedy zákaz cílené reklamy

Mediální manipulace a dezinformace na internetu



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

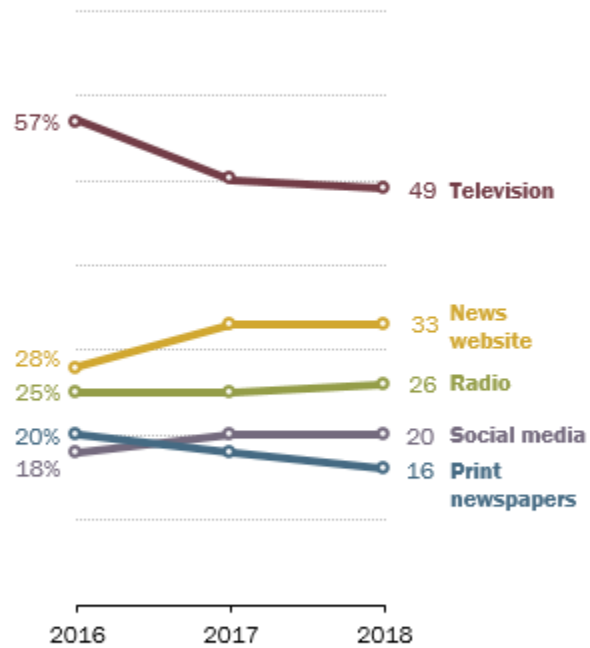
Psychologická podstata

- 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”?

Zdroje informací

More Americans get news often from social media than print newspapers

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



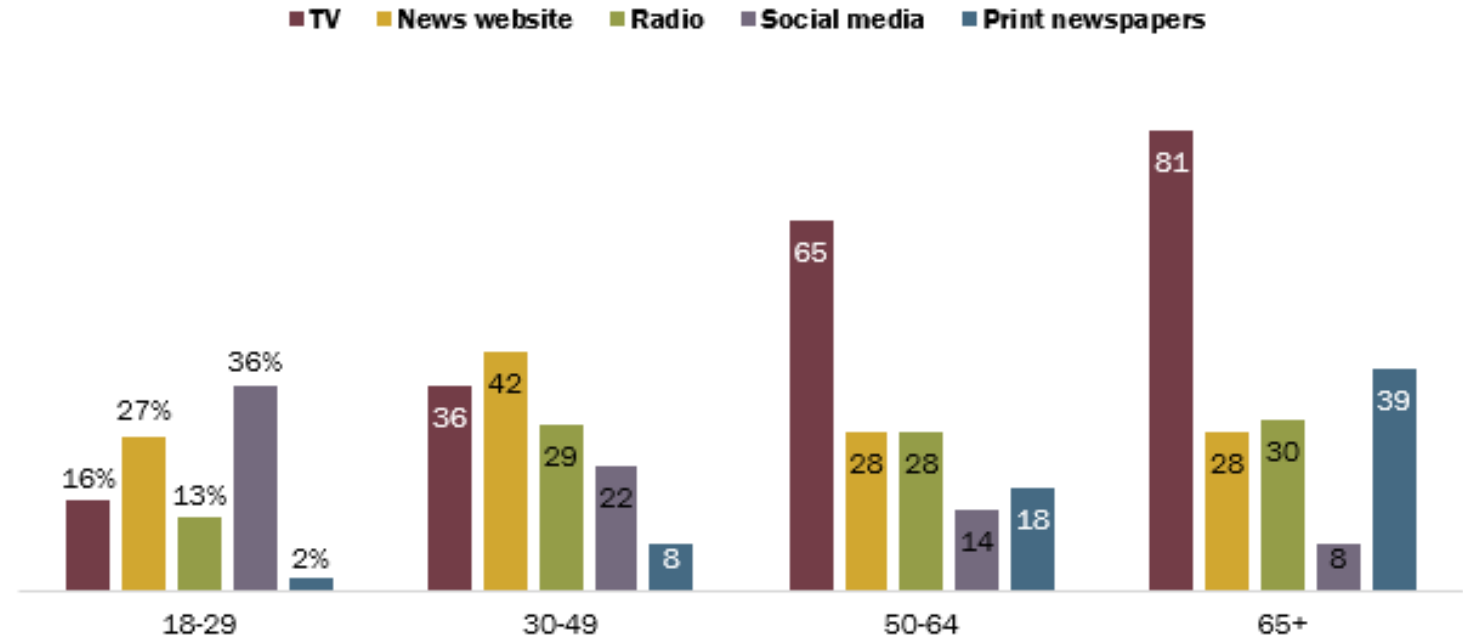
Note: The difference between social media and print newspapers in 2017 was not statistically significant.

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.

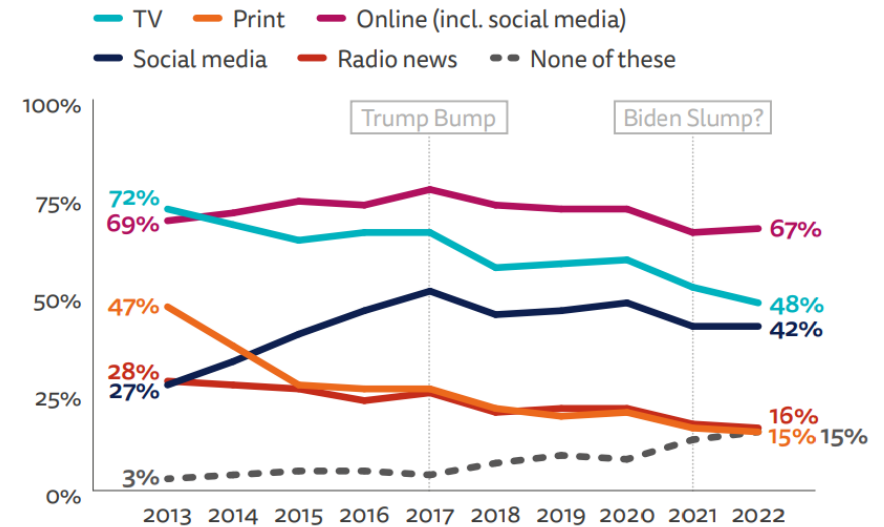
PEW RESEARCH CENTER

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

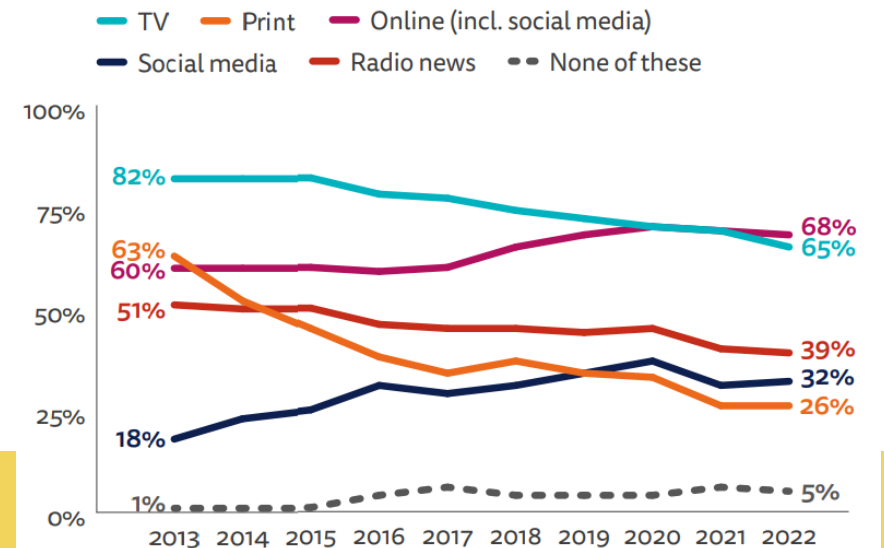
Zdroje informací

- Výrazná proměna za posledních 20 let
- Pokles TV a tištěných novin
- Vzestup online zpráv a sociálních médií
 - Personalizovaný obsah
 - Více náchylné na šíření dezinformací

PROPORTION THAT USED EACH SOURCE OF NEWS IN THE LAST WEEK (2013-22) - USA



PROPORTION THAT USED EACH SOURCE OF NEWS IN THE LAST WEEK (2013-22) - GERMANY



Zpravodajské servery



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

Ekonomika zpravodajských serverů

- Mnoho autorů, omezené zdroje
- Autoři musí naplnit kvóty na délku obsahu a čtenost
 - Kvalita je mnohdy druhořadá
- Investigativní žurnalistika je drahá
 - Rozhovory, ověřování informací
- Poctivou investigativní žurnalistiku čte málo lidí
- Tlak na zisk → Tlak na čtenost zpráv
- Tlak na tvorbu „levných“ zpráv
 - Lze napsat od stolu s využitím internetu
 - Místo ověřování stačí odkaz nebo screenshot
 - důvěryhodnost řeší málokdo
- Pokrytí témat jenom proto, že už jsou pokryta jinde
 - Pokud je tato zpráva čtená u konkurence, je potřeba o ní také psát

Nestrannost vs. Objektivita

- Extrémní pojetí nestrannosti: Všechny názory dostávají stejný prostor
- Včetně nepravdivých, manipulativních, či nehumánních
- Vede k šíření názorů, které vůbec nestojí za pozornost
 - Legitimizace extremismu
- Svoboda slova jako argument pro šíření rasismu, ponižování žen, ...
 - Rozlišení, zda něco **mohu** říct vs. zda **bych měl**
 - Obava z obvinění z cenzury
- Přitom cílem zpravodajství je předávat a vysvětlovat (ověřená) **fakta**
- **Objektivita** má více kritérií:
 - Faktická správnost (relevance, informativnost, pravda)
 - Nestrannost (vyváženost, neutralita)

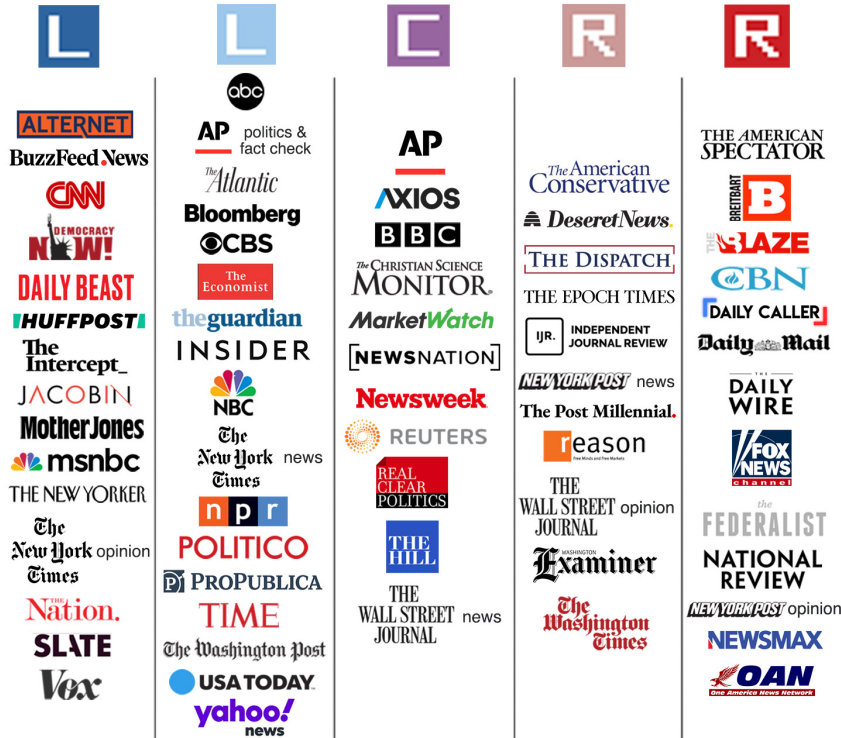
Maximalizace užitku

- 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

Media Bias Charts

AllSides™ Media Bias Chart™

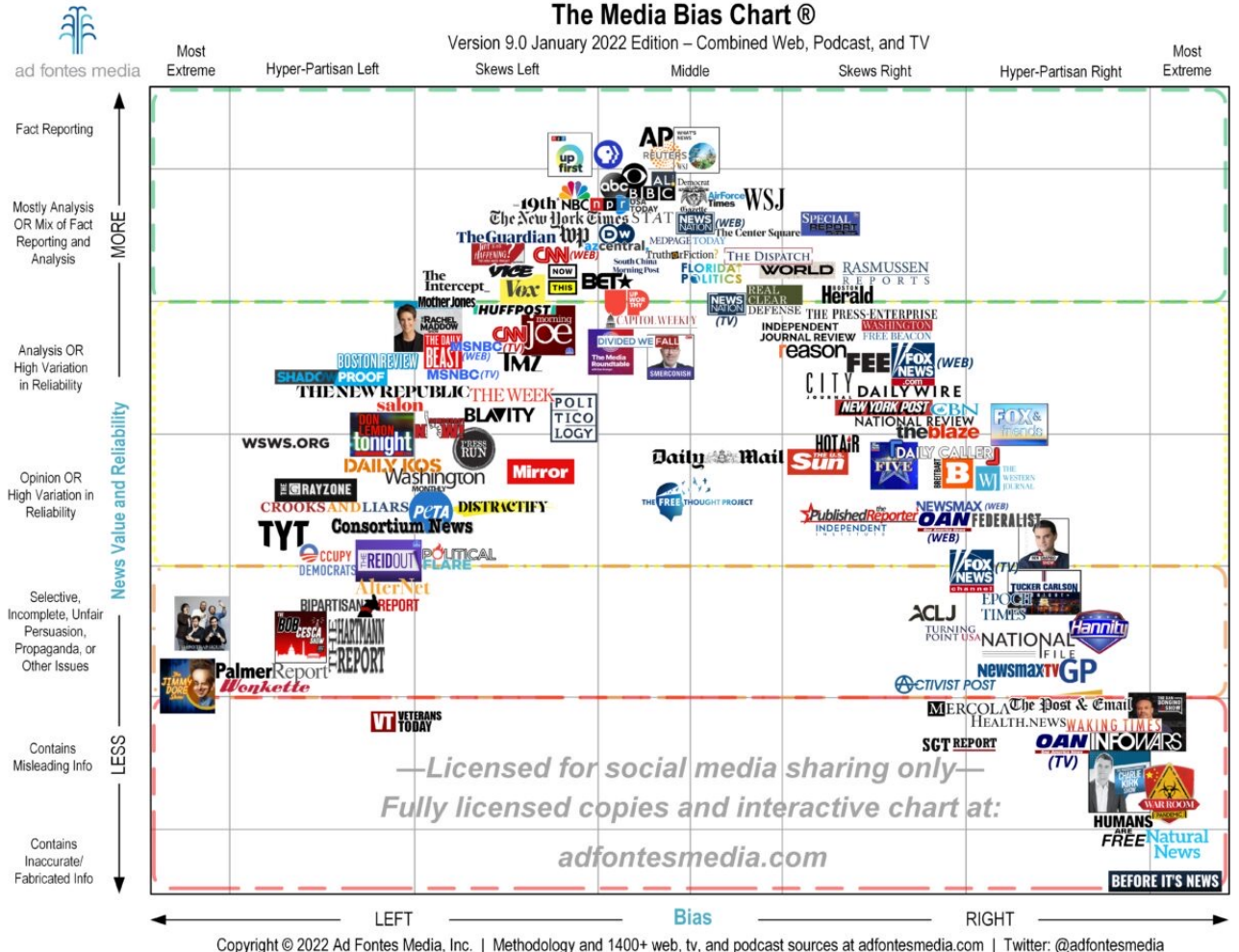
All ratings are based on online content only — not TV, print, or radio content. Ratings do not reflect accuracy or credibility; they reflect perspective only.



L LEFT L LEAN LEFT C CENTER R LEAN RIGHT R RIGHT

AllSides Media Bias Ratings™ are based on multi-partisan, scientific analysis. Visit AllSides.com to view hundreds of media bias ratings.

Version 6 | AllSides 2022



—Licensed for social media sharing only—
Fully licensed copies and interactive chart at:

adfontesmedia.com

Copyright © 2022 Ad Fontes Media, Inc. | Methodology and 1400+ web, tv, and podcast sources at adfontesmedia.com | Twitter: @adfontesmedia

Metody mediální manipulace

- Ředění informací
 - A odvedení pozornosti jinam
- Zdůrazňování nepodstatných problémů
 - K odvedení pozornosti od reálných (vážných problémů)
- Vytvoření umělého problému
 - A nabídka jeho řešení
- Využívání citů a pudů
 - Zejména strach (z neznámého)
- Útok na elity
 - Každý přece “selským rozumem” pozná...

Manipulace volbou slov

- Practicing **pro-life** litigators know that Trump judges are saving lives by permitting restrictions on abortion to go into effect.

<https://thefederalist.com/2020/04/24/david-french-needs-to-stop-slandering-trump-supporting-christians/>

- Bojovník za svobodu vs. Terorista
- Uprchlíci / Migranti / Přistěhovalci
- Válka / Invaze / Agrese / Speciální vojenská operace

REPORT: OVER 80 MILLION CHILDREN AT RISK AS CORONAVIRUS DISRUPTS VACCINATION SCHEDULES



by JOSHUA CAPLAN 22 May 2020 970



Tens of millions of children under 12 months are potentially at risk for diseases such as diphtheria and polio as the **Chinese** coronavirus pandemic interrupts routine vaccinations, according to data published by global public health experts on Friday.

Dezinformace na sociálních médiích



Dezinformace na sociálních 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

Příklad: PizzaGate

- Konspirační teorie šířená před prezidentskými volbami v USA
- Uniklé e-maily Johna Podesta, šéfa kampaně Hillary Clinton údajně obsahovaly kódované zprávy
 - Představitelé Demokratické strany a několik restaurací v USA
 - Měly provozovat obchod s lidmi a nabízet sex s dětmi
- Šíření konspirace přes Twitter, 4chan, 8chan...
- Zaměstnanci dotčených restaurací dostávali výhružky smrtí
 - V několika případech se v restauracích i střílelo
- Konspiraci uvěřilo 46 % voličů Donalda Trumpa a 17 % voličů Hillary Clinton
- Více viz https://en.wikipedia.org/wiki/Pizzagate_conspiracy_theory

Dilemata

- Novináři: Psát o tom, jaké zprávy se šíří?
- Uživatelé sociálních sítí: Sdílet?

- Psát, sdílet?
 - A zároveň kritizovat, vyvracet, vzdělávat
 - Ale šířit původní zprávy mezi více lidí

- Ignorovat
 - Tvářit se, že neexistují
 - Neposkytnout oponentní názor
 - Čekat na něco horšího, co už nebude možné ignorovat

- Co je správné?

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

“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

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

Automatická detekce dezinformací



Automatická 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
- Tradiční metody zkoumající text nestačí
- Využití dodatečných informací
 - Ale kde je vzít, když se jedná o aktuální zprávy?
- Zohlednění uživatelských profilů
 - Ale jak?

Detekce dezinformací: Formulace problému

- Článek
 - Atributy vydavatele
 - Atributy obsahu (nadpis, text, obrázek)
- Šíření
 - Uživatel U vytvoří příspěvek P v čase T
 - Šíření je množina trojic (U, P, T)
 - Příspěvek P typicky obsahuje (odkaz na) článek \check{C} i s jeho atributy
- Detekce dezinformací je binární klasifikační problém
 - Vstup: množina šíření
 - Výstup: Rozhodnutí, zda se jedná o dezinformaci či nikoliv

Fáze 1: Feature extraction

- Obsah zprávy: Autor, titulek, text, obraz nebo video
 - Lingvistické rysy – senzační (clickbait) titulek, struktura vět, přesvědčovací jazyk
 - Grafické rysy – analýza obrazu (nesourodé části ukazující na úpravy fotografie)
- Sociální kontext
 - Uživatelé – detekce botů a kyborgů
 - Příspěvky a reakce na ně – pochybnosti, úžas
 - Sociální síť – vztahy mezi uživateli

Fáze 2: Vytvoření modelu

- 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

Datasey

- BuzzFeedNews
 - 1627 článků od devíti agentur
 - Publikované mezi 19. - 27. 9. 2016
 - Fakta ověřena novináři BuzzFeedu
- LIAR
 - 12 836 tvrzení oštkovaných lidmi
 - Štítky hodnotící pravdivost (5bodová stupnice)
- BS Detector
 - Rozšíření do prohlížeče
 - Vyhodnocuje důvěryhodnost odkazů vedoucích z článků
 - Původní množina domén oštkována manuálně
- CREDBANK
 - 60 milionů tweetů o 1000 událostech
 - Ohodnoceno 30 placenými anotátory

Servery na ověřování faktů

- <https://www.factcheck.org/>
- <https://hoax.cz/cze/>
- <https://manipulatori.cz/>
- <https://cesti-elfove.cz/>

Shrnutí

- Naivní realismus a potvrzovací zkreslení
- Stále více lidí čerpá informace z internetu
 - Zpravodajské servery → Ekonomický tlak na levné a čtené zprávy
 - Sociální média → Sociální bubliny
- Dezinformace se šíří 3x rychleji než pravdivé zprávy
- Detekce dezinformací je stále otevřený problém
 - Provozovatelé sociálních médií nemají motivaci je řešit
 - Analýza obsahu (jazykové rysy dezinformací)
 - Analýza šíření sociální sítí

Zdroje

- Phillips, W. (2018). The Oxygen of Amplification. *Data & Society*, 22
 - https://datasociety.net/wp-content/uploads/2018/05/2-PART-2_Oxygen_of_Amplification_DS.pdf
- Shu et al., 2017: Fake News Detection on Social Media: A Data Mining Perspective
 - <https://dl.acm.org/doi/abs/10.1145/3137597.3137600>

Dilemma Game



Dilema: Psát pro své publikum?

Svoji závěrečnou práci zpracovávám pro jistou vládní organizaci. Když jim prezentuji své závěry, nelíbí se jim, že jsou příliš vágní a nebude z nich tedy možné učinit jasné politické prohlášení. Organizace mne požádá, abych závěr své práce formuloval jasněji a určitěji. Nicméně na základě dat, která mám, není možné tak jasné závěry učinit. Můj vedoucí práce mi poradí, abych se naučil psát pro své publikum, a že bych měl být schopen odvážných tvrzení. Navíc je naděje, že tato organizace bude financovat návazné projekty, na nichž bych se mohl podílet.

- A. Přepíšu závěry tak jak po mě organizace chce.
- B. Odmítnu závěr přepsat.
- C. Nechám práci tak, jak je, ale přidám k ní Shrnutí, v němž budou závěry jasnější.
- D. Požádám staršího profesora, o němž vím, že velmi dbá na dodržování vědeckých postupů, aby v této věci rozhodnul.

Esej

Studuji informatiku na vysoké škole a v jednom dobrovolném teoretickém předmětu mám za úkol odevzdat zápočtovou esej. Od starších studentů vím, že vyučující, který daný předmět vede, je velmi benevolentní a pro úspěšné ukončení stačí poslat i prázdný dokument, který nepůjde otevřít. Vyučující totiž vždy kontroluje jen to, jestli studenti něco poslali, nekontroluje však obsah dokumentu. Mám poslední 2 týdny na předání eseje, přičemž čas na napsání eseje by se dal efektivně využít na jiné předměty, které končí zkouškou. Co mám dělat?

- A. Rozhodnu se poslat prázdný dokument a nepodniknu žádné další kroky.
- B. Rychle napíšu esej nízké kvality a odevzdám ji, díky čemuž nebudu mít výčitky svědomí za zneužití benevolence učitele, a zároveň budu mít dost času na důležité zkoušky. Nepodniknu žádné další kroky.
- C. Rozhodnu se poctivě napsat esej a nepodniknu žádné další kroky.
- D. Přístup učitele se mi nelíbí, protože neplní své pracovní a pedagogické povinnosti, a proto se rozhodnu daný problém formálně nahlásit vedení fakulty.
- E. Rozhodnu se jít za daným učitelem a upozorním ho na nevhodnost jeho jednání a vyzvu ho, aby si plnil poctivě své povinnosti. Počítám s tím, že taková výzva ho nepotěší a může mít pro mě následky.

Příští týden

- Příští týden (16. 3.) přednáška odpadá!
- Analýza Chat GPT ve zvoleném předmětu (20 bodů)
 - Zkuste vyřešit zadané úkoly pomocí Chat GPT
 - Kriticky zhodnoťte užitečnost Chat GPT pro řešení úkolů
 - Identifikujte hrozby a příležitosti
 - Navrhněte způsob, jak smysluplně využít Chat GPT
- Termín: 31. března
 - Za každý den zpoždění penalizace 2 body
 - Odevzdávárna v IS

Úkoly na předášku 23. 3.

- Téma: Filtrování informací a cenzura
 - Kdy lze filtrování informací považovat za etické?
- Přečíst si článek
 - Should You Have The Right To Be Forgotten On Google? Nationally, Yes. Globally, No.
 - https://www.huffpost.com/entry/google-right-to-be-forgotten_b_6624626
- Podívat se na video
 - A Guide to Everything Banned in China
 - <https://www.youtube.com/watch?v=wohoKgnWcBA>