

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
 - Terminologie, nejčastější zdroje informací
 - Ekonomika zpravodajských serverů
 - Šíření dezinformací na sociálních sítích
- Detekce dezinformací
- Téma: Filtrování informací a cenzura
 - Zákony proti šíření dezinformací
 - Nová Velká čínská zeď
 - Právo být zapomenut
 - Blokování dezinformačních webů
 - Filtrování v univerzitní síti
- Dilema: Kontrola dat sesbíraných pro závěrečnou práci

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

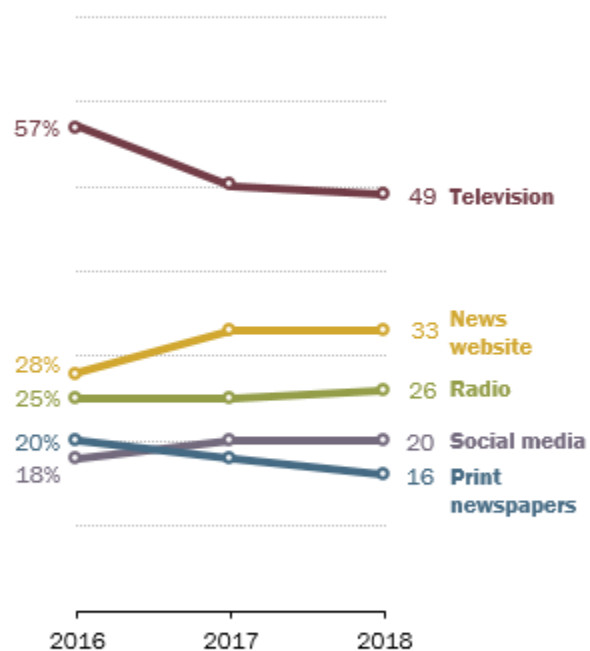
Opakování: Psychologická podstata

- 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



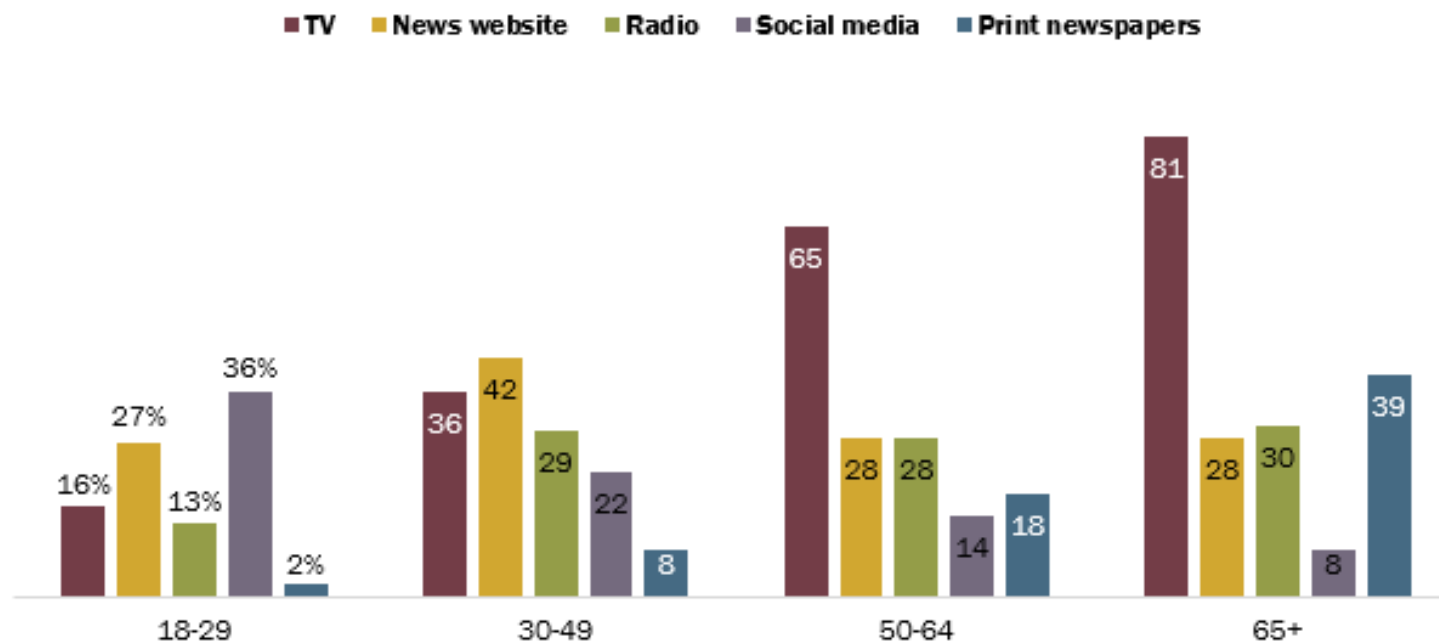
Note: The difference between social media and print newspapers in 2017 was not statistically significant.

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

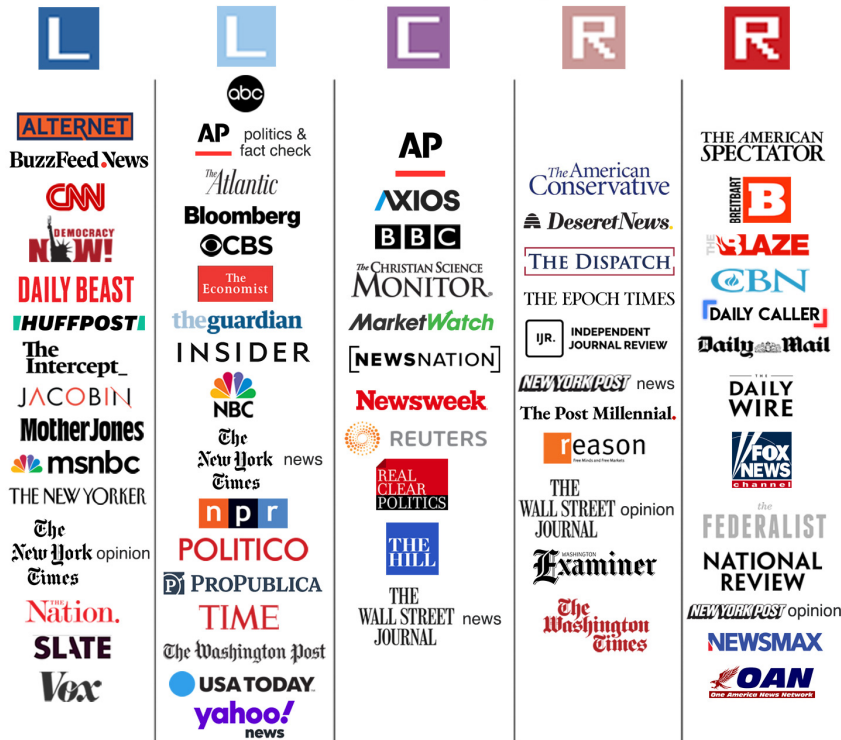
Opakování: Maximalizace užitku

- 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: Media Bias Charts

AllSides™ Media Bias Chart™

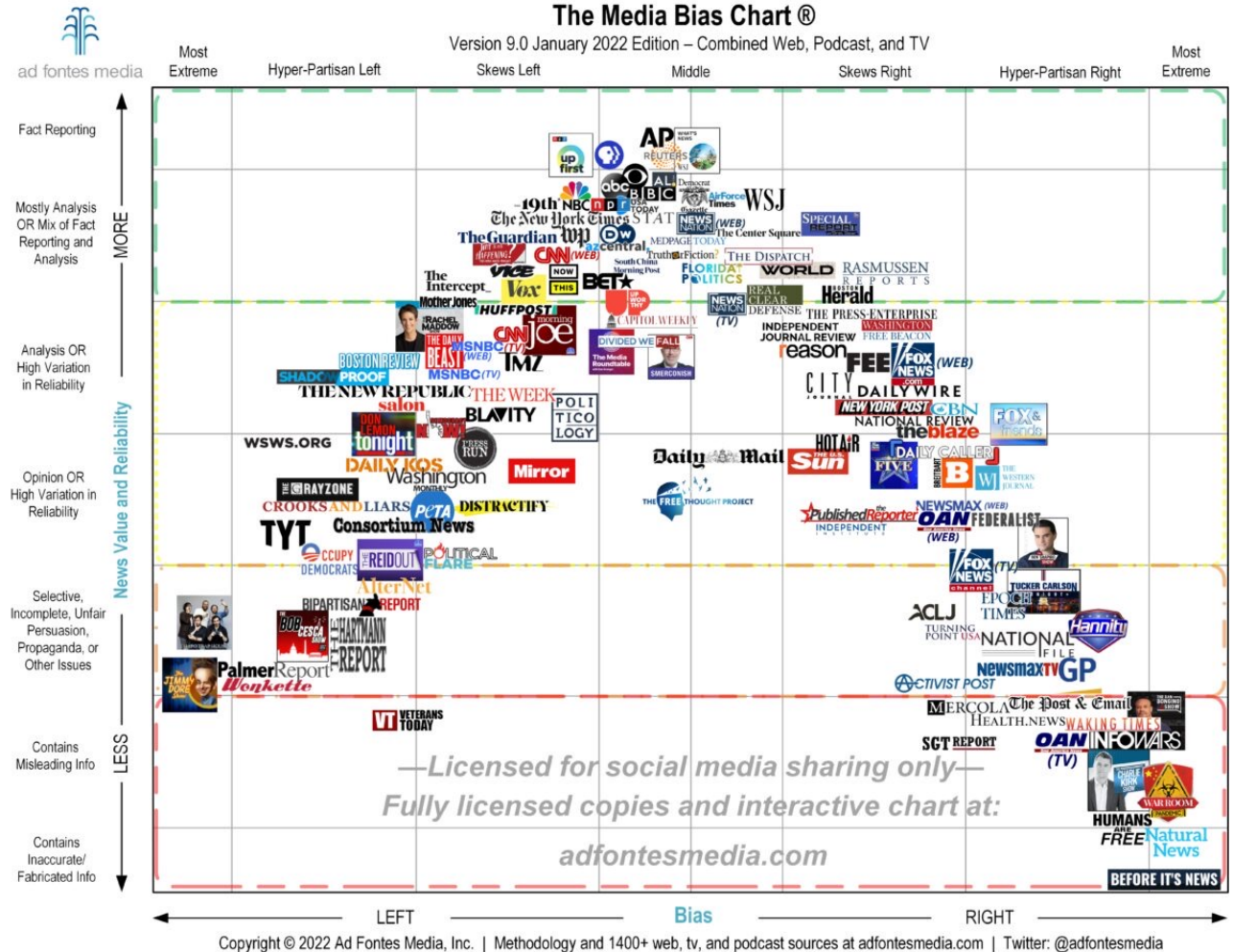
All ratings are based on online content only — not TV, print, or radio content. Ratings do not reflect accuracy or credibility; they reflect perspective only.



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



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: Opakování: 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 vyhružky 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

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

**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 oštitkovaných lidmi
 - Štítky hodnotící pravdivost (5bodová stupnice)
- BS Detector
 - Rozšíření do prohlížeče
 - Vyhodnocuje důvěryhodnost odkazů vedoucích z článků
 - Původní množina domén oštitkována 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

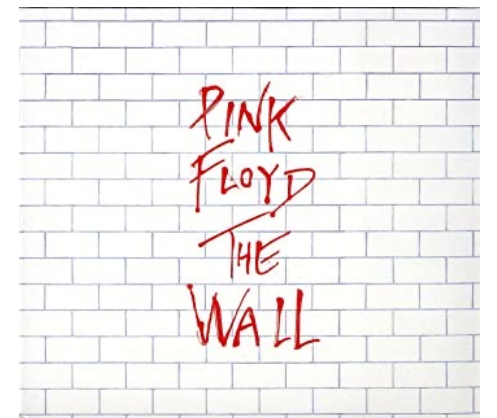
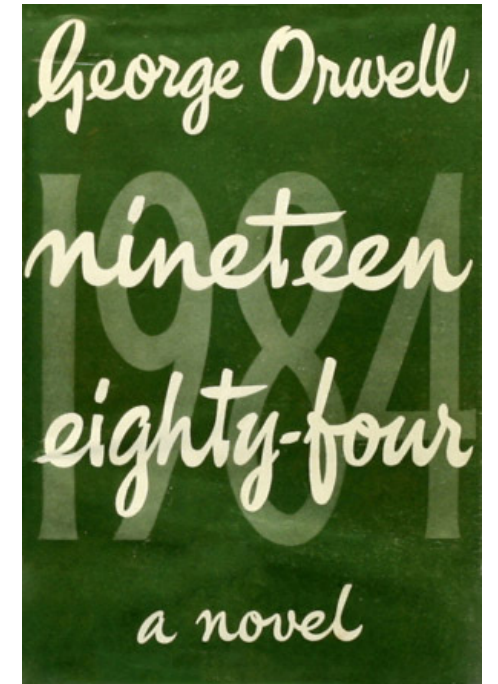
- Phillips, W. (2018). The Oxygen of Amplification. *Data & Society*, 22
 - https://datasociety.net/wp-content/uploads/2018/05/2-PART-2_Oxygen_of_Amplification_DS.pdf
- Shu et al., 2017: Fake News Detection on Social Media: A Data Mining Perspective
 - <https://dl.acm.org/doi/abs/10.1145/3137597.3137600>

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

Blokování dezinformačních webů

- 25. února (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?

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
- 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.

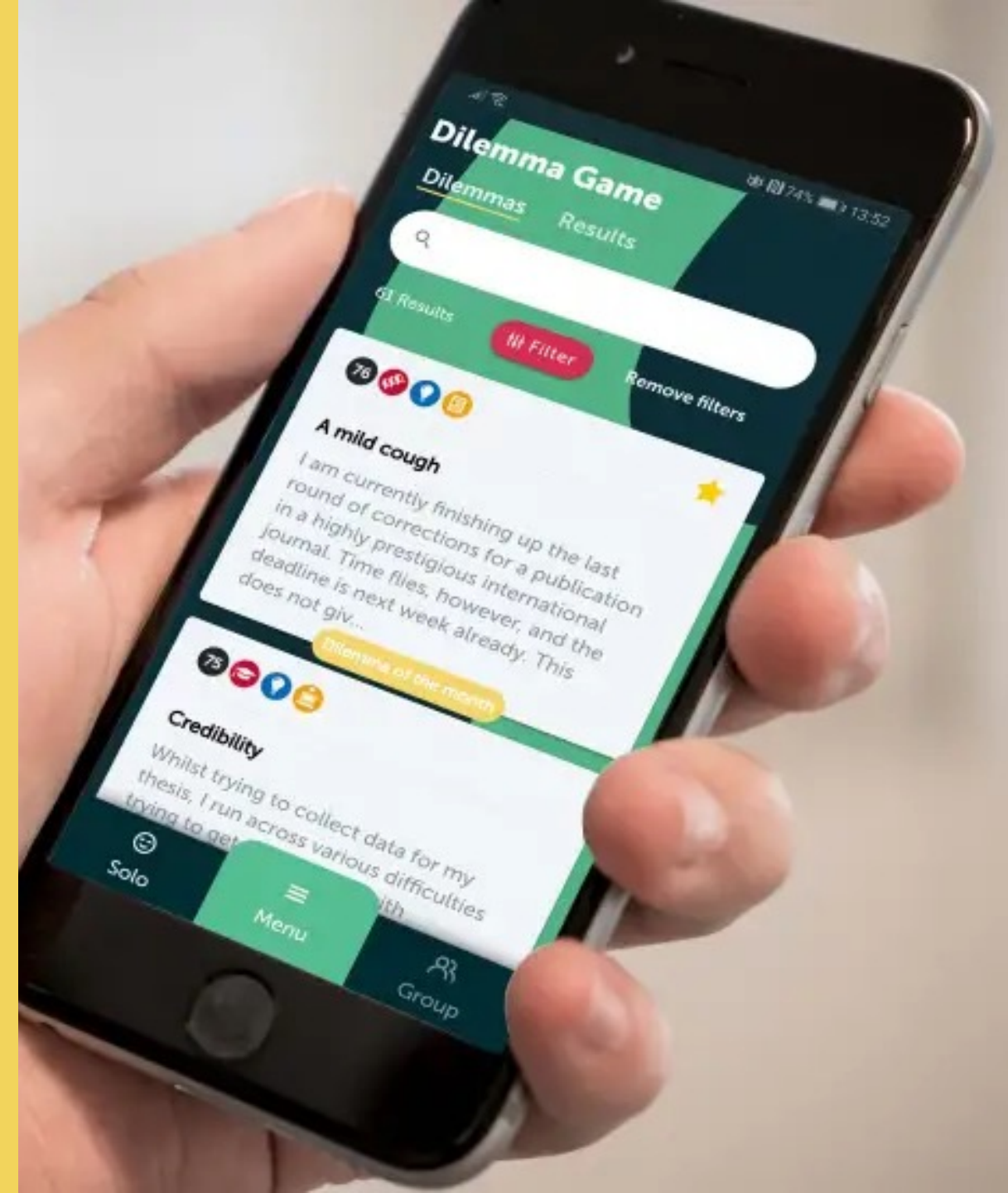
Internet usage is routinely monitored and logged.

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.

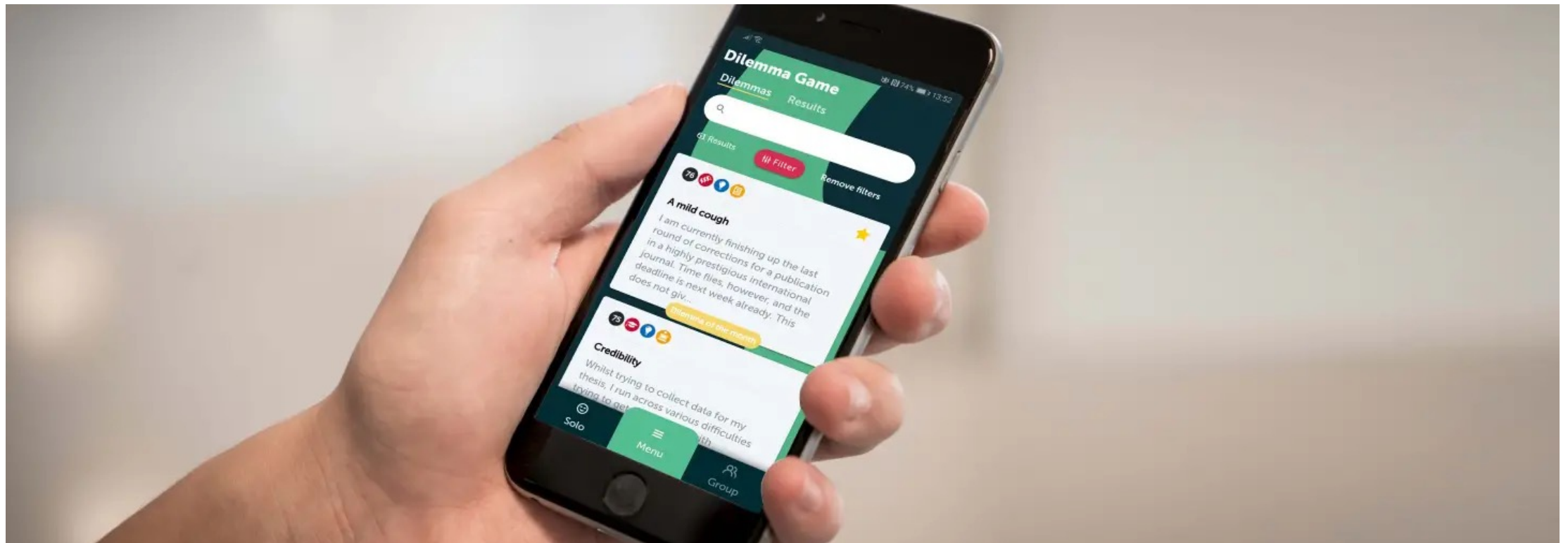
Note that permission for this access will have to be authorised by your Head of Department or your manager. Massey University's Internet usage policies for staff and students can be found [here](#).

Dilemma Game



Dilemma Game

- Doporučuji nainstalovat aplikaci
- <https://www.eur.nl/en/about-eur/policy-and-regulations/integrity/research-integrity/dilemma-game>



Dilema: Kontrola dat

- Pro svoji závěrečnou práci jste sesbírali data od respondentů, kterým jste slíbili anonymitu. Tato data máte uložena na fakultním serveru.
 - Na jiné fakultě se objevil případ fabrikace dat v diplomové práci, což vyústilo v plošnou kontrolu dat ve studentských pracích na celé univerzitě.
 - Byli jste požádáni o informace o vašich respondentech, aby se potvrdilo, že jste skutečně sesbírali data od reálných osob.
- A. Anonymita respondentů je klíčová. Žádné informace nikomu neposkytnu.
 - B. Dodám informace o identitě respondentů způsobem, který neumožní spárování respondentů s daty o jejich osobě.
 - C. Poskytnu plný přístup ke svým datům pod podmínkou, že kontrolující osoba podepíše závazek mlčenlivosti.
 - D. Poskytnu plný přístup ke svým datům. Institucionální kontrola má vyšší prioritu než mnou daný slib anonymity.

Příští týden

- Zvaná přednáška
 - Mgr. Ing. Jiří Marek, ÚVT MU
- Téma: Internet a autorská práva
 - „Remixová“ kultura
 - Volné dílo a licence Creative commons
- Úkoly
 - Přečíst si první dvě kapitoly knihy "Made with Creative Commons"
<https://creativecommons.org/use-remix/made-with-cc/>
 - Zběžně se podívat na <https://creativecommons.org/>