

**MUNI**  
**FI**

# **Etické nakládání s daty II**

Mgr. Tomáš Foltýnek, Ph.D.

[foltynec@fi.muni.cz](mailto:foltynec@fi.muni.cz)

# Osnova dnešní přednášky

- Opakování: Etika internetového výzkumu
- Organizačně: Témata esejů a prezentací
- Dilema: Ukončit spolupráci s ruskými institucemi?
  
- eHealth a trasování covid-19
  - Možnosti deanonymizace anonymních dat
- Kapitalismus dohledu
  
- Dilema: Kontrola dat sesbíraných pro závěrečnou práci

**OPAKOVÁNÍ:  
Etika  
internetového  
výzkumu**



# Opakování: Výzkum nad daty Dropboxu

- Článek “[Was it Ethical for Dropbox to Share Customer Data with Scientists?](#)”
  - Anonymizovaná data 16 000 výzkumníků z 1000 kateder/ústavů/oddělení
  - Informace o složkách (sdílení, využívání)
  - Informace o vědeckém výkonu (počet článků, počet citací)
- Cíl [studie](#): Zjistit, jak spolupracují nejlepší týmy
  - Pracují v malých týmech (průměr 2,3 členů, vs. 3,0 u nejhorších)
  - Pracují na projektech déle (průměr 172 dnů vs. 130 u nejhorších)
  - Udržují stabilní týmy (průměr 5 projektů vs. 3,5 u nejhorších)
  - Rovnoměrné rozdělení práce mezi členy týmu (nikdo se „neveze“)
  - Zkušenější pracují více (63 % vs. 58 % u nejhorších)
- Bylo etické data sdílet?

# Opakování: 3T Dataset

- Tastes, Ties and Time
- 2008: [Výzkumníci z Harvardu](#) publikovali [dataset](#) 1700 Facebookových profilů
  - Profily patřily studentům z Harvardu; anonymizované, kódované
  - Ukazuje vývoj sociálních sítí studentů v průběhu studia
  - Veřejně dostupná jen kódová kniha; dataset na vyžádání
- Obrovský potenciál pro sociální vědce
- Problémy
  - Nikdo ze studentů nedal ke sběru dat souhlas
  - Demografická data, obor a další údaje umožňovaly deanonymizaci
    - A tedy zveřejnění informací určených jen pro přátele
  - Některé údaje byly přístupné jen uživatelům ze stejné školy
    - Jejich zveřejnění by tedy porušilo uživatelská nastavení
- Dataset nakonec nebyl zveřejněn
- Podrobný rozbor viz Zimmer, M. (2010). “But the data is already public”: on the ethics of research in Facebook. *Ethics and Information Technology*, 12(4), 313–325. <https://doi.org/10.1007/s10676-010-9227-5>

# Opakování: OkCupid Dataset

- 2016 dánští výzkumníci [Kirkegaard & Bjerrekaer publikovali dataset](#) uživatelů seznamovací stránky OkCupid
  - Data získali pomocí web scraperu
  - Uživatelské jméno, věk, pohlaví, sexuální orientace, město
  - Odpovědi na otázky o náboženství, politických názorech, drogy, nevěra, svazování během sexu...
- Všechny údaje byly dobrovolně vloženy uživateli s vědomím, že budou veřejně dostupné
  - Výzkumníci je poskytli v „užitečnější formě“
- Výzkumníci neměli ke sběru dat souhlas platformy, uživatelů, ani etické komise univerzity

# Opakování: GDPR

- **G**eneral **D**ata **P**rotection **R**egulation
- Právní rámec EU pro ochranu osobních údajů
- Osobní údaj = každý údaj týkající se **identifikované nebo identifikovatelné** osoby
- Pseudonymizace (kódování) nemusí zamezit identifikovatelnosti!
- **Anonymní** informace pro statistické nebo výzkumné účely nepodléhají GDPR

# Opakování: Zásady zpracování os. údajů

- Zákonnost – musí existovat důvod pro zpracovávání dat
- Transparentnost – subjekt musí vědět, jaká data se uchovávají
- Omezení účelu – data nesmí být využívána v rozporu s tímto účelem
- Omezení uložení – uložení ve formě umožňující identifikaci
- Minimalizace – uložení jen nezbytných dat
- Integrita, přesnost, důvěrnost



# Opakování: Otázky pro nakládání s daty

- 10 otázek podle Data Science Process Alliance
- Jaké **zákony** a předpisy se vztahují na daný projekt?
- Jak víme, že je etické data využít k danému **účelu**?
- Kdo je zodpovědný za **minimalizaci etických rizik**?
- Jak se využití dat **dotkne práv** jednotlivců?
- Jak se **propojování dat** dotkne soukromí a anonymity jednotlivců?
- Jak víme, že data jsou k danému účelu **vhodná**?
- Jak rozpoznat a **snížit zkreslení** v datech?
- Jak **zmírnit subjektivitu** při návrhu systému?
- Jak velká **transparentnost** je potřeba a jak ji dosáhnout?
- Jak pravděpodobný je **mylný výklad výsledků** a jak mu zabránit?

# Opakování: Norimberský kodex (1947)

- První kodifikace **etických zásad výzkumu na lidech**
  - Reakce na výzkumy prováděné nacisty během 2. světové války
1. **Dobrovolný souhlas** lidské osoby je absolutně nevyhnutelný.
  2. Pokus by měl být takový, aby přinesl **plodné výsledky** pro dobro společnosti, nebyl zjistitelný jinými metodami nebo studijními prostředky a neměl by být ve své podstatě náhodný a zbytečný.
  3. Pokus by měl být navržen a založen na výsledcích experimentů na zvířatech a na znalosti přirozené historie onemocnění nebo jiného studovaného problému tak, aby **očekávané výsledky ospravedlňovaly provádění pokusu**.
  4. Pokus by měl být prováděn tak, aby **se zabránilo veškerému zbytečnému tělesnému a duševnímu utrpení a zranění**.
  5. Neměl by být prováděn **žádný pokus, pokud lze předpokládat, že dojde ke smrti nebo poranění**. Snad kromě pokusů, kde experimentující lékaři slouží jako pokusné objekty.
  6. **Stupeň rizika**, které je třeba podstoupit, by nikdy neměl překročit míru určenou **humanitárním významem problému**, který má být pokusem vyřešen.
  7. Měly by být provedeny vhodné přípravy a zajištěno náležité **vybavení na ochranu účastníků** pokusu proti sebemenší možnosti zranění, zdravotního postižení nebo smrti.
  8. Pokus by měly provádět pouze **vědecky kvalifikované osoby**. Měl by být vyžadován nejvyšší stupeň dovedností a péče při všech fázích pokusu u všech účastníků pokusu.
  9. Během pokusu by měl mít účastník **možnost pokus ukončit**, pokud se dostal do takového tělesného nebo duševního stavu, kdy mu pokračování v pokusu přijde nemožné.
  10. Odpovědný **vědec musí být připravený ukončit pokus** v kterékoli fázi, má-li, na základě svých nejlepších zkušeností, důvod věřit, že by pokračování v pokusu mohlo mít za následek zranění, postižení nebo smrt pokusného objektu.

# Opakování: Belmontská zpráva (1978)

- Reakce na „Tuskegee Syphylis Study v USA “ (1972)
  - Zkoumání průběhu syfilitidy na afroameričanech, kterým nebyl podán penicilin
- Základní principy
  - **Úcta k lidem**: Respekt lidské autonomie, důstojnost účastníků výzkumu
    - Anonymita, důvěrnost
  - **Užitečnost**: Maximalizovat přínosy výzkumu, minimalizovat rizika pro účastníky, požadavek neškodit
  - **Spravedlivost**: férovost při výběru subjektů, rovnost
- Vodítko pro posuzování výzkumu etickými komisemi
- Základní principy převzaté i do evropských a dalších předpisů
  - Ženevská deklarace (1947), Helsinská deklarace (1964),...

# Opakování: Etika internetového výzkumu

- **Je výzkum nad daty sociálních sítí výzkumem na lidech?**
- Živý jedinec se stává subjektem výzkumu, pokud výzkumník
  - Shromažďuje informace nebo biologické vzorky pomocí intervence či interakce
    - Interakce zahrnuje komunikaci nebo jiný kontakt mezi výzkumníkem a subjektem výzkumu
  - Získává, využívá, studuje, analyzuje či generuje identifikovatelné soukromé informace či identifikovatelné biologické vzorky
- **Soukromá informace**
  - lze důvodně očekávat, že nebyla vůbec zaznamenána, nebo
  - byla poskytnuta pro stanovený účel a lze důvodně očekávat, že nebude zveřejněna
- **Různé výklady „důvodného očekávání“, „stanoveného účelu“ a „zveřejnění“**
  - Více viz např. <https://plato.stanford.edu/entries/ethics-internet-research/>
- **Etika internetového výzkumu je relativně mladá**
  - Nejsou konkrétní závazné předpisy; Aplikují se obecnější normy

**Dotazy?**

**Zajímavosti?**



# Organizačně

- Témata esejů?
- Témata a termíny prezentací na přednáškách
  
- Prezentace na přednášce – očekávání za 20 bodů
  - Zajímavé a relevantní téma
  - Prezentace 10 – 15 minut
  - Formulování konkrétních etických dilemat
  - Diskuse 10 – 15 minut

# Morální dilema

- V reakci na válku na Ukrajině
  - Je správné ukončit spolupráci s ruskou institucí / firmou / výzk. týmem?
  - Je správné odmítnout vědecký článek jenom proto, že autor je z Ruska?
- Izolovat Rusko jako celek?
- Podporovat protiválečné ostrůvky v Rusku?
- Co je správné?
- Co má větší potenciál zastavit válku?
- Kde je hranice?



Anna Abalkina  
@AbalkinaAnna

Negotiations between UA and RU are taking place now. The head of the Russian delegation is Medinsky, Putin's advisor and ex-minister of culture. Just to give you an idea who he is I attach the print screens of his PhD theses. Each colored cell(except grey)=page=plagiarism detected

Medinsky Vladimir Rostislavovich  
OF WORLD DEVELOPMENT AND THE PROBLEMS OF FORMING THE FOREIGN POLICY STRATEGY OF RUSSIA IN THE CONDITIONS OF  
Dissertation for the degree of doctor of political sciences  
Specialty 23.00.04

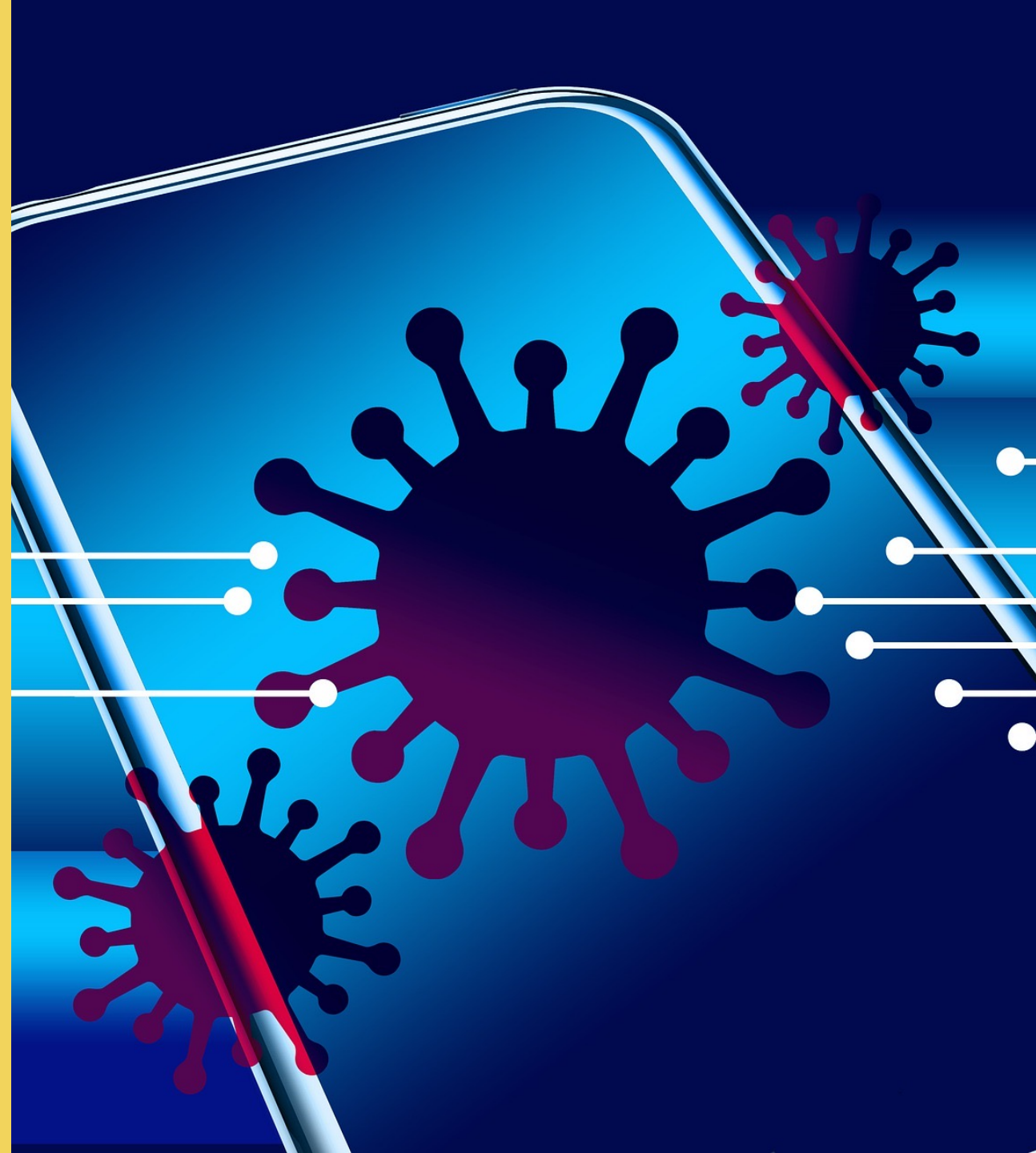
Russian Federation RAN/EPA  
Medinsky Vladimir Rostislavovich (1999). Borrowing table

Page	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	
41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	
81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	
161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	
201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	
241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	
281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	
321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	

Publications, Tables, Figures - are not subject to textual analysis  
Special problems of forming the foreign policy of the Russian Federation in a changing world: dissertation (Advocatum and Eurasianism), M., 1994.  
Mirovaya ekonomika i mezhdunarodnye otnosheniya, No. 10, 1996.

are not subject to textual analysis  
Development and the problems of formation of Russia's foreign policy: Dissertation, 1997.  
Foreign Policy and International Law: Translation from German and editing by O. STREKAL (Grim) and A. ZOTOVA (Dobrowa), 1998. Pp. 67-72. Professor Christiane Gernert - official page: <https://www.tu-berlin.de/~gernert/>

**eHealth**  
**Trasování covid-19**





# eHealth: Aplikace na prevenci nemocí

- Aplikace na vyhodnocování příznaků: Měření teploty, typ kašle
- Náramky, hodinky a aplikace na sběr dat (teplota, tep, tlak)
  - Obrovské datasety pro medicínský výzkum
  - Prevence nemocí, včasné varování
- Prevence nemocí přenášených komářími kousnutím
  - Malárie, abroviry (dengue, zika, žlutá zimnice...)
  - 2021: [Rešerše](#) výzkumníků z univerzity ve Freiburgu
    - Zkoumali studie o přínosech mobilních aplikací na prevenci abrovirálních nemocí
    - Nejčastější způsob (25 z 36 aplikací) je **sledování uživatelů**
      - Reportování nemoci → Předpovídání ohnisek nákazy
      - Sledování polohy → Nepřiblížili se k místům, kde se množí nakažení komáři?
    - Druhý nejčastější způsob je vzdělávání (mLearning)

# Trasovací aplikace na covid-19

- Velký rozmach během první vlny (jaro 2020)
- Přehledová studie (březen 2021):
  - 120 aplikací v 71 zemích
  - Jen v USA 23 různých aplikací
  - Nejvíce masového sledování v Evropě
  - 73 aplikací na zdravotní certifikáty
    - 60 z nich nemá řádné Zásady ochrany osobních údajů
- Využité technologie
  - GPS tracking: Nepřesné, energeticky náročné, “kontakt” přes zed’
  - Bluetooth: Často nefunkční, obcházení API → narušení zabezpečení
  - Manuální trasování (ruční zadávání kontaktů do formuláře)

# ”Anonymizovaná” geolokace

- [Studie](#) (2020) ukázala, že z anonymizovaných dat o poloze je možné získat údaje o bydlišti a zaměstnání s 98% přesností
  - Rada: Šifrovat ID uživatelů pomocí soli, která se několikrát denně mění → Rozbití sekvencí polohových dat
- 2013: New York City Taxi & Limousine Commission
  - Zveřejnění datasetu o 173 milionech jednotlivých jízd
    - Čas a místo začátku a konce, jízdné, spropitné
    - ID taxikářů hashována
  - Anonymita datasetu byla jen zdánlivá (viz [článek v Big Data & Society](#))
    - Deanonymizace taxikářů → Informace o jejich příjmech
    - Spojení s dalšími zdroji (blogy, články,...) → Sledování celebrit
    - Odvození adres lidí navštěvujících strip kluby

# Důležité parametry trasovacích aplikací

- Míra přijetí aplikace
  - V USA bylo ochotno nainstalovat aplikaci méně než 50 % lidí
  - V Německu ochota dosáhla cca 50 % lidí (viz [studie v Nature Human Behav.](#))
    - Z toho 1/3 by se přidala až poté, co by dostali zaplaceno 1, 2 nebo 5 EUR
  - K úspěšnému fungování aplikace je třeba min. 60 % (spíše 80 %)
- Důvěra obyvatel
  - V závažnost problému, který aplikace řeší
  - V instituci vydávající aplikaci
  - Ve funkčnost aplikace
  - V ochranu osobních údajů
- Funkčnost aplikace
- Provázanost s dalšími prvky systému → Aplikace sama nic neřeší

# Doporučení DCEI

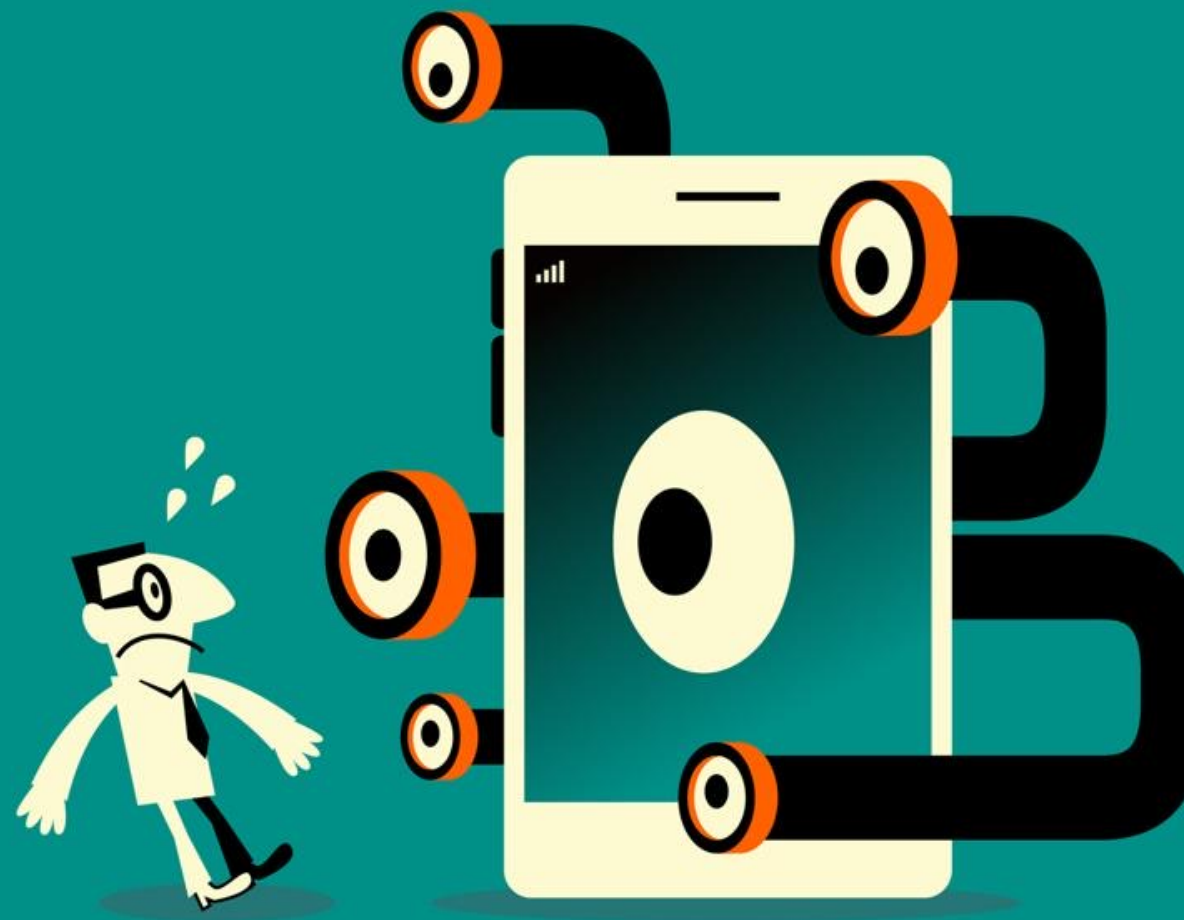
- Centre for Data Ethics and Innovation, UK
- [Soubor doporučení pro trasovací aplikace \(2020\)](#)

<b>Value</b>	There must be good reason to <b>believe that that the app will provide sufficient net-value</b> back to the citizen or society as a whole so as to justify its introduction and any adverse consequences for individuals. There must be a compelling value proposition to users.
<b>Security (and privacy)</b>	Data sharing and storage should be secure. The <b>data collected should be minimised and protected</b> as much as possible, so <b>users' privacy is preserved</b> .
<b>Accountability</b>	There must be a <b>reliable and understandable decision-making process</b> in place to manage the app - with clear accountability, particularly with regards to introducing new functionality, data collection or use cases.
<b>Transparency</b>	A strong and demonstrable <b>commitment to transparency enables scrutiny</b> and helps to build trust.
<b>Control</b>	Users should be able to <b>see what kinds of data are held</b> about them so that they can understand how it is impacting on decisions.

# Taiwan: Příklad funkčního trasování

- Studie v International Journal of Infectious Diseases
  - Popisuje jen první vlnu v roce 2020
  - 487 případů nákazy, 8051 blízkých kontaktů
- Centrální systém trasování
  - Kontakt s nakaženým → 2 týdny v karanténě
  - Trasování do tří dnů (než se člověk stal infekční)
  - Sebereportování → cca polovina záznamů v databázi
- Trasování + roušky + rozestupy → Pokles  $R_0$  z 2,5 na cca 0,7
  - Taiwan se zbavil covidu **bez lockdownu**
- Co (a proč) fungovalo na Taiwanu a nefungovalo v ČR?

# Kapitalismus dohledu



# Shoshana Zuboff: Surveillance capitalism

- Český “kapitalismus dohledu”
- 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
- Posun paradigmatu
  - Hledáme na Googlu → Google hledá v nás
  - Využíváme sociální sítě → Sociální sítě využívají nás
  - Zásady ochrany osobních údajů → Zásady sledování (dohledu)
- Nejde však jen o pasivní využití sesbíraných údajů...



# Výzkum “nakažlivosti” nálady

- 2012 výzkum Facebook + Cornell University
- 2014 publikován v [Proc. of the National Academy of Sciences of the USA](#)
- Experiment na 689 003 uživatelích
  - Vynechání citově zabarvených slov ze zpravodajského kanálu ovlivňuje náladu
    - Vynechání pozitivních slov → Sdílení spíše negativního obsahu
    - Vynechání negativních slov → Sdílení spíše pozitivního obsahu
  - Prokázání masivního šíření dobré/špatné nálady v sociální síti
  - Důkaz, že “nakažlivost nálady” nevyžaduje přímou sociální interakci
- Výzkum proveden bez souhlasu uživatelů a v rozporu s podmínkami používání služby
- Facebook se omluvil za “špatnou komunikaci” experimentu
- Do podmínek používání byla přidána část týkající se výzkumu

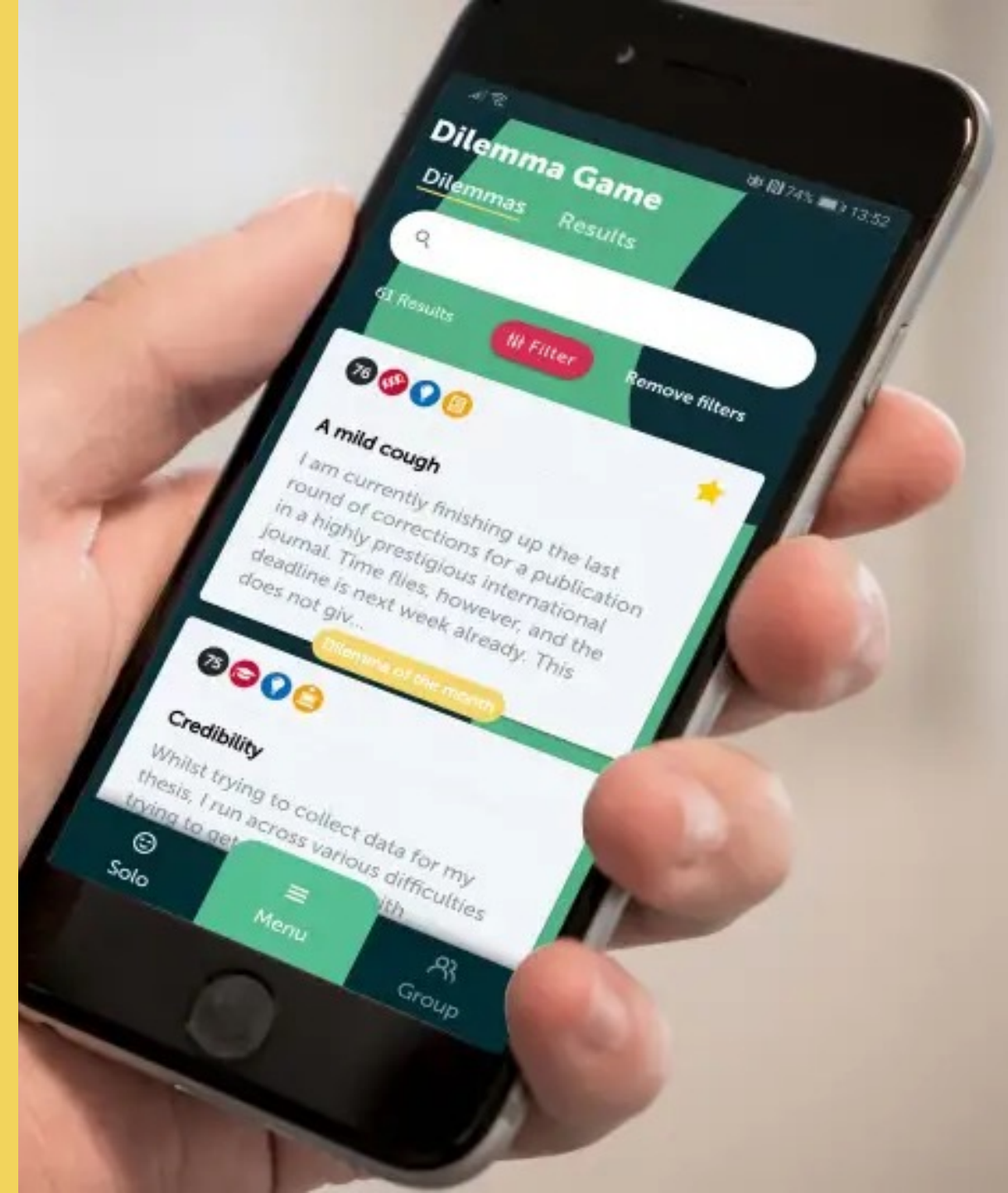
# Skandál Cambridge Analytica

- 2013: Aplikace “This is your digital life”
  - Dotazník k vytvoření psychologického profilu uživatele
  - Přihlášení přes Facebook
  - Aplikace měla přístup k osobním údajům uživatelů a k údajům o přátelích
- 2013 – 2014 vyplnilo 270 tis. uživatelů
  - Aplikace získala přístup k údajům o 87 milionech uživatelů
    - Většina z nich nikdy s poskytnutím údajů nesouhlasila
- Využití získaných údajů pro cílený politický marketing
  - Např. kampaně Donalda Trumpa, Brexit,...
- Odhalení v roce 2018
  - Cambridge Analytica zbankrotovala
  - Facebook zaplatil pokutu 500 000 GBP za neoprávněné poskytnutí dat

# Shoshana Zuboff: Surveillance capitalism

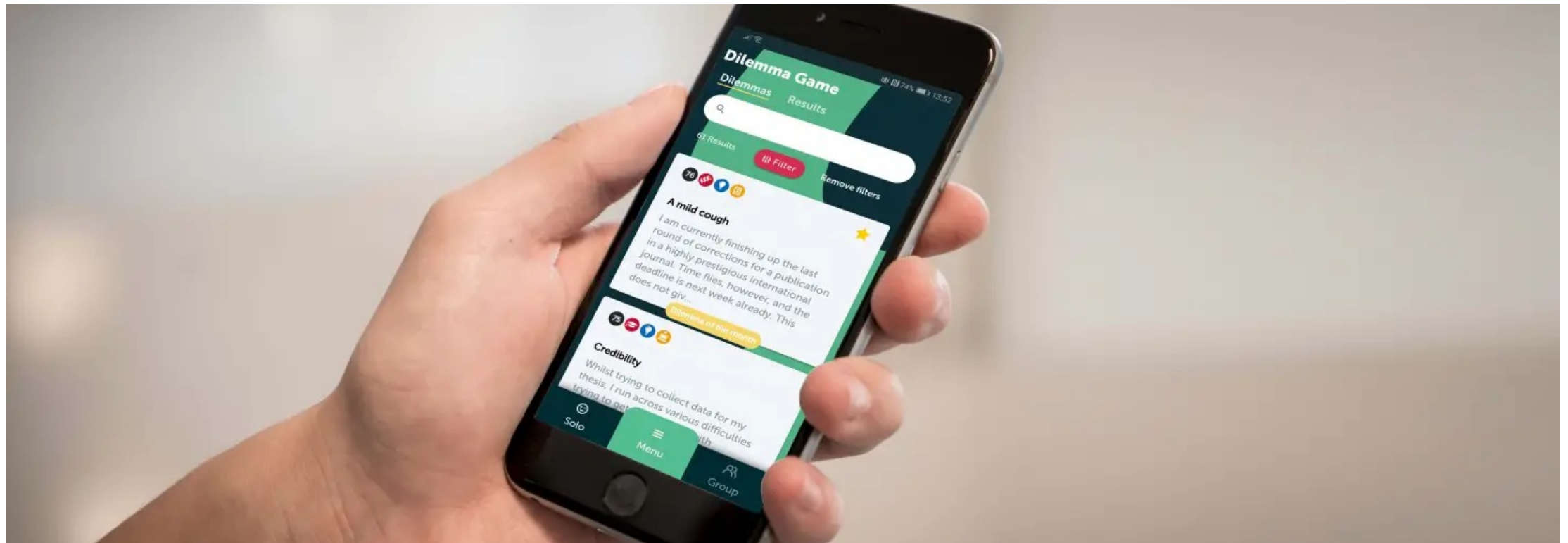
- Pasivní využívání našeho nitra → Jeho aktivní ovlivňování
  - Nálada uživatelů, na co kliknout, co si koupit, koho volit...
- Za účelem zisku dochází k “dálkovému ovládnutí lidí”
  - Ohrožení lidské svobody a demokracie
- Řešení: Regulace
  - Kapitalismus je regulován – antimonopolní zákony, ochrana zaměstnanců
  - Vybrané segmenty trhu jsou nelegální – obchod s lidmi, s orgány,...
  - Stejně tak by mělo být nelegální obchodování s lidskou budoucností
- Souhlasíte?

# 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.

# Úkoly na příště

- Příští týden přednáška odpadá; Další přednáška bude 16. 3.
- Podívat se na film Sociální dilema (The Social Dilemma)
- Přečíst si článek
  - Annette Zimmermann, Bendert Zevenbergen
  - AI Ethics: Seven Traps
  - <https://freedom-to-tinker.com/2019/03/25/ai-ethics-seven-trps/>
- Téma: Zkreslení informací v algoritmech
  - Etické aspekty transformace dat na informace a znalosti
  - Etické aspekty prezentace dat