Dr. Roberto Yus
https://robertoyus.com/

Data Privacy

L01 - Data Privacy and Regulations

Icons from https://thenounproject.com/

Defining Privacy

- Wikipedia → *“ability of an individual or group to seclude themselves or information about themselves, and thereby express themselves selectively”*

- Merriam-Webster → *“freedom from unauthorized intrusion” , “the quality or state of being apart from company or observation”*

Defining Privacy

- Private information sometimes equated to **personal data** / **personally identifiable information (PII)**
  - *“(1) any **information that can be used to distinguish or trace an individual's identity**, such as name, social security number, date and place of birth, mother's maiden name, or biometric records;*
  - *(2) any other **information that is linked or linkable to an individual**, such as medical, educational, financial, and employment information." [1]*

What should be considered private?

[1] NIST: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-122.pdf

Privacy Principles

- **Minimization**: what you collect, who has access to it, and how long you keep it
- **Choice**: user's option to share information
- **Access**: user's right to review, correct and possibly delete the data you hold
- **Transparency**: disclosure of what, how and with whom data will be shared

Privacy vs Security

- While privacy and security go hand in hand, they are not the same concepts
- Security is about protecting both the information you choose to share and that you choose not to share from getting into the hands of others

Can you have security without privacy?

What about privacy without security?

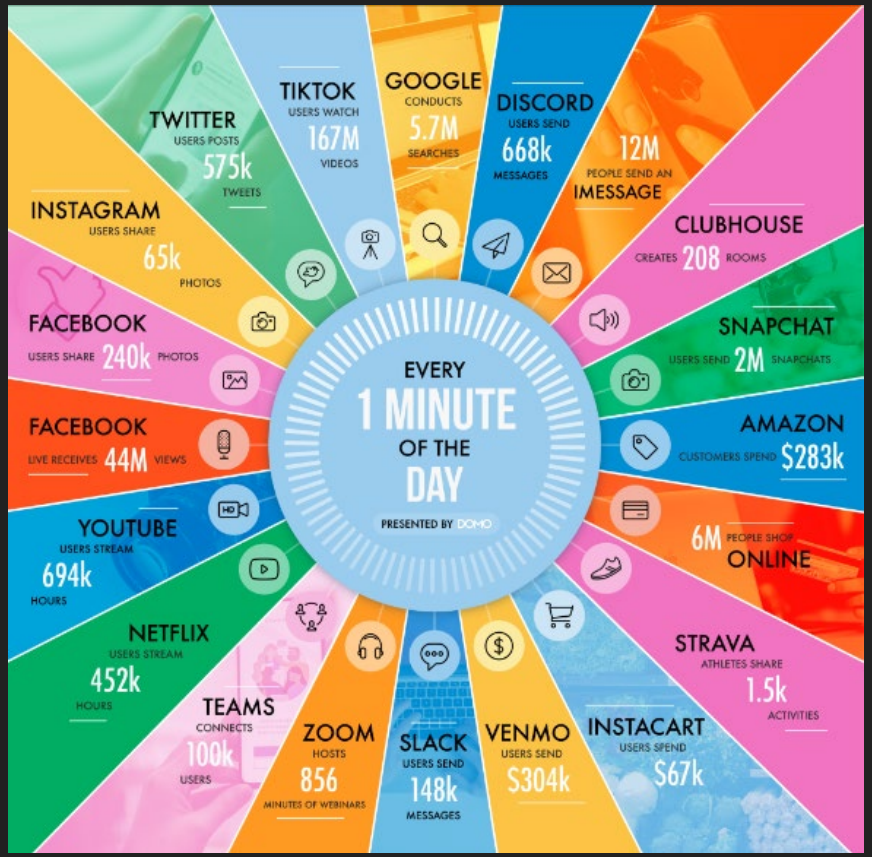
DuckDuckGo ✓
@DuckDuckGo

Replying to @DuckDuckGo

Security without privacy is like a house made of bullet-proof glass. No-one's getting inside but your personal life is still on display.

11:16 AM · Jul 26, 2017 · Twitter Web Client

Today's Privacy Challenge



Source: https://www.domo.com/learn/infographic/data-never-sleeps-9

#	Company	Image recognition		
		Face recognition	Environment recognition	Product recognition
1	Facebook	•	•	•
2	Instagram	•	•	•
3	Tinder	•	•	
4	Grindr	•	•	
5	Uber	•		
6	TikTok	•	•	•
7	Strava			
8	Spotify			
9	Myfitnesspal			
10	Clubhouse	•		
11	Credit Karma			
12	Twitter	•	•	
13	Airbnb	•	•	
14	Lidl Plus			•
15	American Airlines			
16	eBay			•
17	Sleepcycle			
18	Paypal			
19	Slimming World			
20	Whatsapp			

What can companies tell from image recognition?

The personal data that recognition software helps companies collect from you

Facial recognition
Recognises people and their key attributes

Background recognition
Detects elements in shot, establishes environment

Object recognition
Can identify an object or product within an image

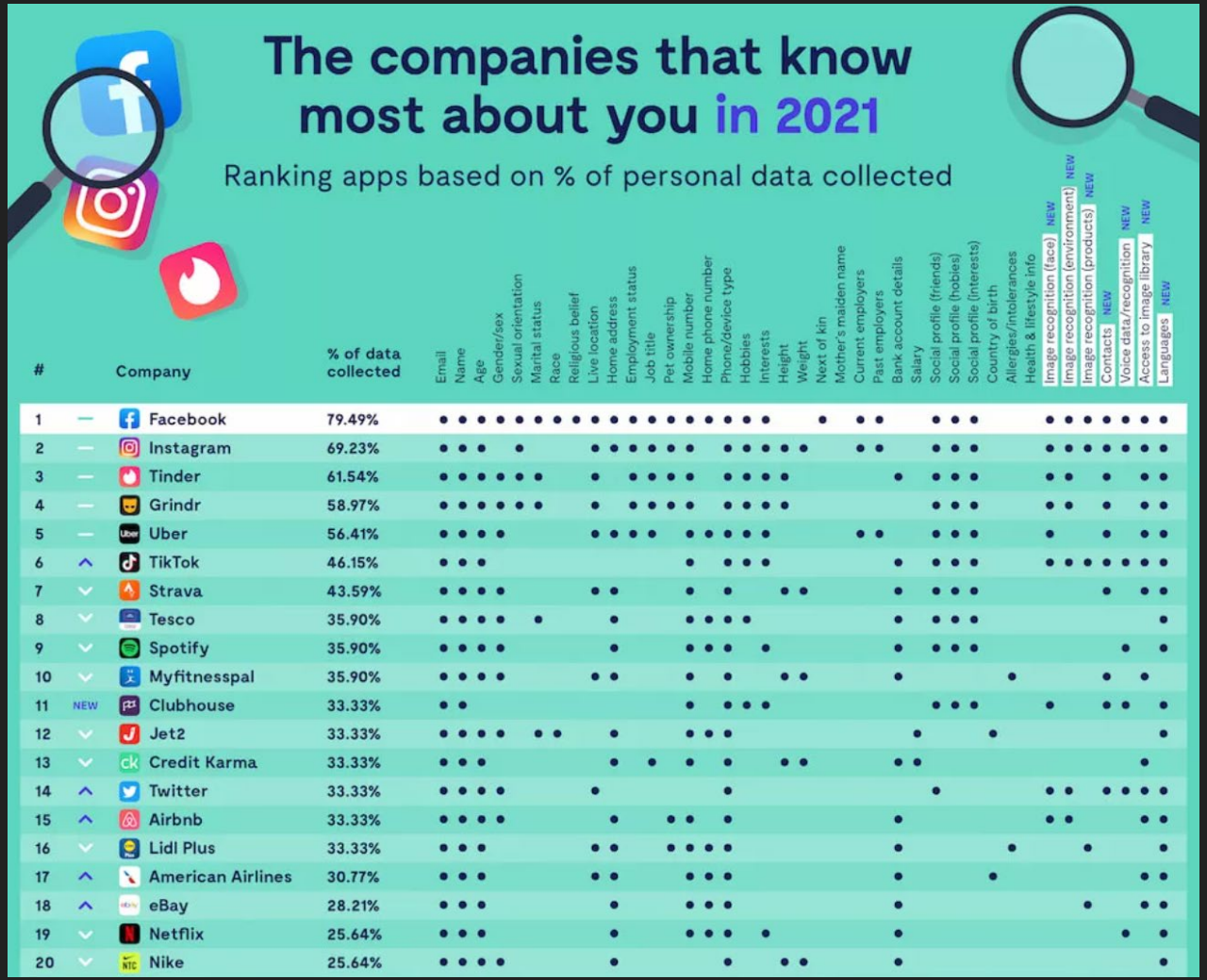
#	Company	Face recognition	Environment recognition	Product recognition	Your contacts	Voice data/ recognition	Access to image library	Languages
1	Facebook	•	•	•	•	•	•	•
2	Instagram	•	•	•	•	•	•	•
3	TikTok	•	•	•	•	•	•	•
4	Twitter	•	•		•	•	•	•
5	Tinder	•	•		•		•	•

Find the full report at **clario.co/blog/which-company-uses-most-data**

clario

The companies that know most about you in 2021

Ranking apps based on % of personal data collected



#		Company	% of data collected
1	—	Facebook	79.49%
2	—	Instagram	69.23%
3	—	Tinder	61.54%
4	—	Grindr	58.97%
5	—	Uber	56.41%
6	^	TikTok	46.15%
7	v	Strava	43.59%
8	v	Tesco	35.90%
9	v	Spotify	35.90%
10	v	Myfitnesspal	35.90%
11	NEW	Clubhouse	33.33%
12	v	Jet2	33.33%
13	v	Credit Karma	33.33%
14	^	Twitter	33.33%
15	^	Airbnb	33.33%
16	v	Lidl Plus	33.33%
17	^	American Airlines	30.77%
18	^	eBay	28.21%
19	v	Netflix	25.64%
20	v	Nike	25.64%

…but, only algorithms access the data…

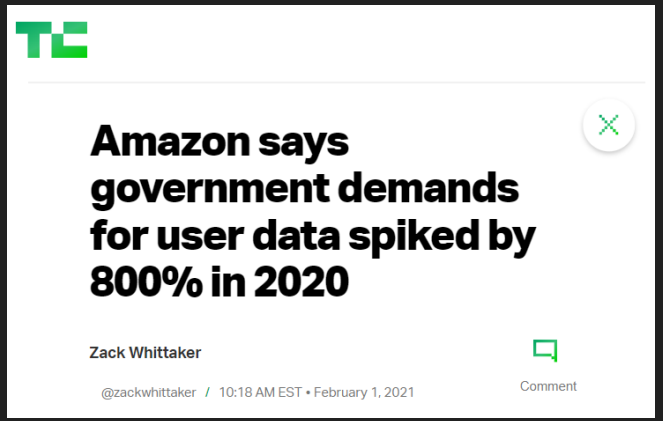
Always?

…but, only algorithms access the data…

Always?

…and they don't share data…

Never?



…but, only algorithms access the data…

Always?

…and they don't share data…

Never?

…but they anonymize data before sharing!

Yes (maybe if it's not a breach)…

Is anonymization enough?

~~Name~~	Age	Condition
~~John Doe~~	59	Cancer
~~Mary Smith~~	78	Covid-19

Case Study: The Netflix Prize

dannypeled.com

- 2007 -- Netflix competition to create the best collaborative filtering algorithm to predict what rating a user would give to a movie based on their previous ratings
- Netflix Prize had a huge positive impact in the research area on recommendations

Case Study: The Netflix Prize

- Dataset: Ratings of almost 500K customers (more than 100M ratings in total)
- Netflix **anonymized** the dataset
  - From their FAQ *“[...] all customer identifying information has been removed; all that remains are ratings and dates.”*

What could go wrong?

Case Study: The Netflix Prize

- Researchers from UT at Austin presented a **statistical de-anonymization attack** using IMDB as background knowledge [1]
- Able to **identify individual users** and determine potentially sensitive information (political views, religious views, or sexual orientation).
- Four Netflix users filled a class action lawsuit.
    - Netflix ended up reaching a settlement with the plaintiffs in 2010.
- Cancellation of projected sequel to the Netflix Prize

[1] Narayanan, A., & Shmatikov, V. (2006). How to break anonymity of the netflix prize dataset. arXiv preprint cs/0610105.

Case Study: AOL Search Data [1]

- Dataset: 20M search queries for 650K users from 2006
- Why released: allow researchers to understand search patterns
- How anonymized: user identifiers removed
  - All searches from same user linked by an arbitrary identifier
- Attacks: many successful attacks identified individual users
  - Ego-surfers: people typed in their own names
  - Zip codes and town names identify an area
  - NY Times identified 4417749 as 62yr old GA widow [2]
- Consequences: CTO resigned, two researchers fired
  - Well-intentioned effort failed due to inadequate anonymization

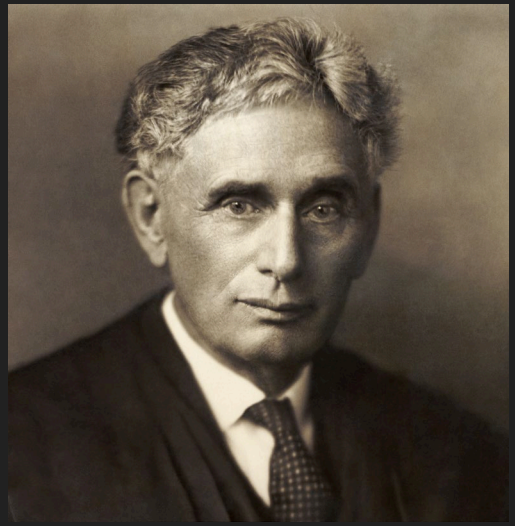
[1] Slide extracted from Cormode, G., & Srivastava, D. (2009, June). Anonymized data: generation, models, usage. SIGMOD (pp. 1015-1018).
[2] https://www.nytimes.com/2006/08/09/technology/09aol.html

Data Privacy Regulations

A Long Time Ago

- The history of privacy rights and regulations starts long before the era of the Internet and Smartphones.
- The Supreme Court has found that the U.S. Constitution (1789) does provide for a right to privacy in its First, Third, Fourth, and Fifth amendments [1]
- 1890, Warren and Brandeis articulated the Right to Privacy [2] *(“right to be let alone”)*: newspapers are the primary sources of *"the unwarranted invasion of individual privacy"* and urged that he courts *"protect the privacy of private life."*

[1] http://law2.umkc.edu/faculty/projects/ftrials/conlaw/rightofprivacy.html
[2] L. Brandeis and S. Warren. The right to privacy. Harvard law review, 4(5), 1890.



*“[the Constitution] conferred, as against the Government, the **right to be let alone** - the most comprehensive of rights and the right most valued by civilized men.“*

Justice Brandeis, dissenting in Olmstead v. US (1928)

Data Protection Regulations

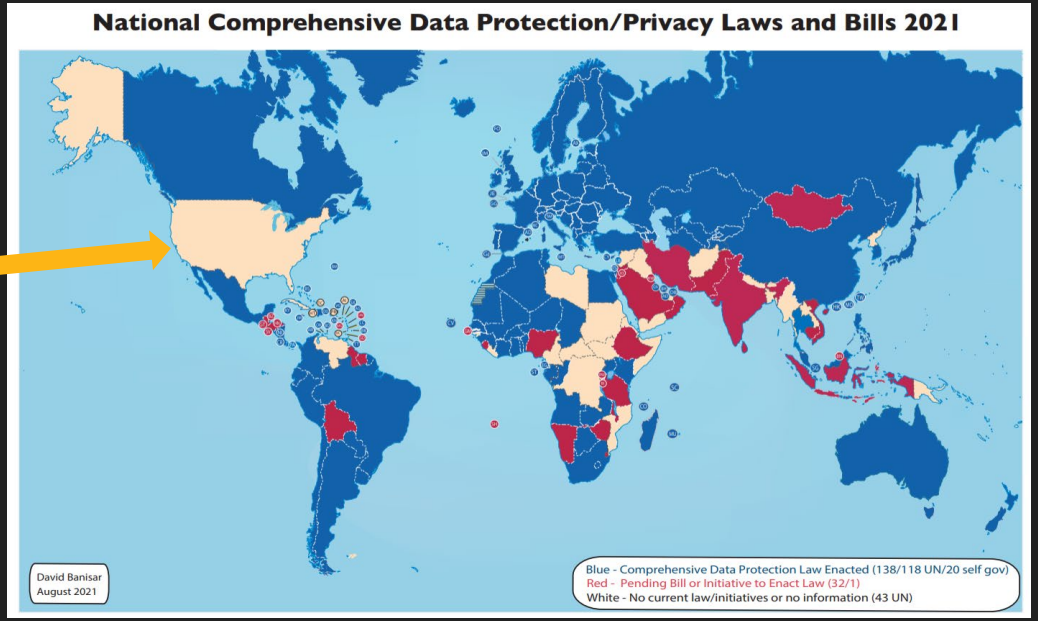
- Technological advances (e.g., computers) enabled governments and corporations to capture and retain large amounts of individuals data since the 1950s.
- Data protection/privacy regulations have been enacted.
- Federal state of Hesse (Germany) passed the first data protection law in the world in 1970 [1].

[1] J. L. Riccardi. The German federal data protection act of 1977: Protecting the right to privacy. BC Int'l & Comp. L. Rev., 6:243, 1983.

Regulations Today (August 2021)

- Over 140 countries, self-governing jurisdictions, territories have adopted national laws and almost 30 countries and jurisdictions have pending bills or initiatives [1]

CCPA in California



[1] National Comprehensive Data Protection/Privacy Laws and Bills 2021, D Banisar - Privacy Laws and Bills, 2021

European General Data Protection Regulation (GDPR)

Slides from Martin Degeling

GDPR

- In effect starting May 25th, 2018
- Unified data protection rules in all 27 member states of the European Union
- Replaced previous Data Protection Directive
  - Unlike Data Protection Directive, GDPR is a law (i.e., other, national laws are not required)
- Is applicable to all services offered within the EU (regardless of where the company is located)

https://gdpr-info.eu/

New and Expanded Rights

- Right to be informed
- Right of access
- Right to rectification
- Right to erasure
- Right to restriction
- Right to data portability
- Right to object
- Right to prevent automated processing, including profiling

Data Protection by Default/Design

- Aka **Privacy by design**

*“Taking into account the **state of the art**, [...] the controller shall [...] implement appropriate technical and organisational measures [...] to implement data-protection principles, such as data **minimisation**, in an effective manner [...].”*

- Collecting data for unknown purposes is forbidden

Data Portability

“The data subject shall have the right to receive the personal data concerning him or her [...] in a structured, commonly used and machine-readable format”

- Right to export data and transfer it to another data controller

Right to be Forgotten

- Aka **Right to Erasure**

“The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay”

- When purpose is fulfilled, consent is withdrawn, unless the controller demonstrates **compelling legitimate grounds** for the processing

Automatic Decision-Making

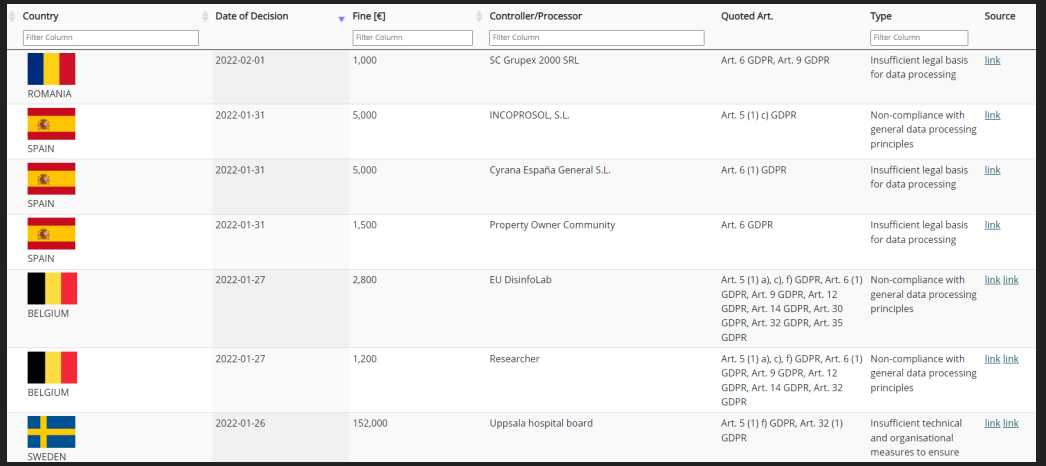
- *“The data subject shall have the right **not to be subject to a decision based solely on automated processing**, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.”*

- Need to identify cases to which this applies and make sure user intervention is possible
- Users have the right to challenge automatic decisions

Regulatory Measures

- **High fines**: up to 20,000,000 EUR or up to 4% of the annual worldwide turnover
- **Data breaches** have to be reported within 72 hours, data subjects have to be notified
- Need for **data protection impact assessment**

https://www.enforcementtracker.com/



Country	Date of Decision	Fine [€]	Controller/Processor	Quoted Art.	Type	Source
ROMANIA	2022-02-01	1,000	SC Grupex 2000 SRL	Art. 6 GDPR, Art. 9 GDPR	Insufficient legal basis for data processing	link
SPAIN	2022-01-31	5,000	INCOPROSOL, S.L.	Art. 5 (1) c) GDPR	Non-compliance with general data processing principles	link
SPAIN	2022-01-31	5,000	Cyrana España General S.L.	Art. 6 (1) GDPR	Insufficient legal basis for data processing	link
SPAIN	2022-01-31	1,500	Property Owner Community	Art. 6 GDPR	Insufficient legal basis for data processing	link
BELGIUM	2022-01-27	2,800	EU DisinfoLab	Art. 5 (1) a), c), f) GDPR, Art. 6 (1) GDPR, Art. 9 GDPR, Art. 12 GDPR, Art. 14 GDPR, Art. 30 GDPR, Art. 32 GDPR, Art. 35 GDPR	Non-compliance with general data processing principles	link link
BELGIUM	2022-01-27	1,200	Researcher	Art. 5 (1) a), c), f) GDPR, Art. 6 (1) GDPR, Art. 9 GDPR, Art. 12 GDPR, Art. 14 GDPR, Art. 32 GDPR	Non-compliance with general data processing principles	link link
SWEDEN	2022-01-26	152,000	Uppsala hospital board	Art. 5 (1) f) GDPR, Art. 32 (1) GDPR	Insufficient technical and organisational measures to ensure	link link

The biggest GDPR fines

1. Amazon —€746 million ($877 million) ...
2. WhatsApp —€225 million ($255 million) ...
3. Google Ireland —€90 million ($102 million) ...
4. 4. Facebook —€60 million ($68 million) ...
5. Google LLC —€60 million ($68 million) ...
6. Google – €50 million ($56.6 million)

https://www.tessian.com/blog/biggest-gdpr-fines-2020/

Discussion

Do we need Data Privacy Regulations?

Are DPRs enough?

Can we make sure that DPRs are updated?

How do we implement DPRs?

…