GDPR, Profiling and Privacy

Engin Bozdag
15 June 2018
Delft Data Science Seminar: Trusted Data Analytics – from bias to trust
About me

Privacy by Design Lead at Philips

Previously:
- PhD student at TU Delft
- Privacy consultant
- General IT

Currently:
- Privacy by Design Lead at Philips
GDPR and Profiling

• Big data and privacy: some examples
• GDPR and general profiling
• Profiling with (semi)automated decision making
• Profiling and automated decision making without human intervention
• Art 29 Working Party Guidelines
Big Data for All: Privacy and User Control in the Age of Analytics

Omer Tene
Haim Striks School of Law

Jules Polonetsky
Future of Privacy Forum
Incremental Effect, Bias, Predictive Analytics

• More personal data, less privacy

• Different conclusions from:
  • “paris”, “Hilton”, and “louvre”
  • “paris”, “Hilton”, and “Nicky”

• Once data is out to public/semi-public sphere => difficult to erase

• Bias in datasets, may consolidate existing prejudices and stereotyping

• Surprise Effect: Target case
Limits of Anonymization

Robust De-anonymization of Large Sparse Datasets

Arvind Narayanan and Vitaly Shmatikov
The University of Texas at Austin

Unique in the Crowd: The privacy bounds of human mobility
Yves-Alexandre de Montjoye, César A. Hidalgo, Michel Verleysen & Vincent D. Blondel
Scientific Reports 3, Article number: 1376 (2013) | Download Citation

On Taxis and Rainbows
Lessons from NYC's improperly anonymized taxi logs

An Adversarial Analysis of the Reidentifiability of the Heritage Health Prize Dataset
Arvind Narayanan
Lack of Access

• The right of access under European Data Protection Directive has been implemented narrowly.

• Organizations provide individuals with little useful information

• Non-compliance was wide around 2017, pre-GDPR
General Data Protection Regulation

Organisations outside the EU that monitor behaviour or offer goods and services to EU residents.

Fines of up to €20 million or 4% global turnover.

Cookies  IP address  Health

Biometric  Genetic
8 Individual Rights

01 Right to be informed
Provide transparency over how personal data are collected, stored, managed, protected, and processed.

02 Right to access
Provide individuals' access to their data and explain how they—and any supplemental data—are used.

03 Reject automated decisions
Comply with requests not to automate decision making using personal data.

04 Right to correction
Correct any personal data if incomplete or inaccurate.

05 Right to be deleted
Remove personal data on request when there is no compelling reason to keep it.

06 Right to restrict processing
Honor requests not to process an individual's data for specific purposes.

07 Right to data portability
Provide copies of all stored data in a portable format.

08 Right to stop processing
Allow individual's data to be stored but not processed.
GDPR Art 22 (1): automated 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.
Profiling

• Automated form of processing
• Carried out on personal data
• objective of the profiling must be:
  • to analyse or predict aspects concerning the individual’s performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements;
• Not every automation is profiling:
  • classify customers based on their age or gender for statistical purposes without making any predictions or drawing any conclusion about an individual.
Legal Effects

- affects someone’s legal rights, such as voting in an election
- affects a person’s legal status or their rights under a contract.
  - cancellation of a contract
  - entitlement to or denial of a particular social benefit granted by law, such as child or housing benefit
  - refused admission to a country or denial of citizenship

"Your recent Amazon purchases, Tweet score and location history makes you 23.5% welcome here."
Similar Significant Effects

- affects the circumstances, behavior or choices of the individuals concerned
- have a prolonged or permanent impact on the data subject
- at its most extreme, lead to the exclusion or discrimination of individuals
  - financial circumstances, such as their eligibility to credit;
  - access to health services
  - Employment opportunities
  - access to education
Example: Automated Decision with or without profiling
Example: Automated Decisions with profiling

Automated CV processing along with psychometric analysis in job recruiting process
How an Applicant Tracking System Reads YOUR RESUME

1. HR RECEIVES YOUR RESUME (along with 100's of others)

2. YOUR RESUME IS RUN THROUGH A PARSER *

* A parser is a computer program that removes styling and breaks down text into recognized strings of characters for further analysis.
3. **The Parser Assigns Meaning to Resume Content**

   - Education
   - Contact Info
   - Skills
   - Work Experience

4. **Employer Uses Keywords to Search Candidates**

   Matching terms are searched from the results collected in step 3.

   **Search Applicants**

   Enter Keywords/Skills
Your resume is scored based on relevancy.

Relevancy takes into account semantic matching of employer search terms and the applicant's years of experience.

If you survive the ATS and score among the highest, you are likely to get called in for an interview.

Resumes **NOT** optimized for Applicant Tracking Systems lack critical keywords and risk never being seen by human eyes.
Example: Automated Decisions with profiling

How Can I Make Sure My Resume Gets Past Resume Robots and into a Human's Hand?

How to persuade a robot that you should get the job

Do more human beings stand a chance against software that claims to reveal what a real-life face-to-face chat can't?

How to beat automated resume screening - Workopolis Blog

How is your Resume Ready for Automated Screening? | Resume Hacking
www.resumehacking.com/ready-for-automated-resume-screening
Jan 2, 2016 - A lot of job seekers seem to be brushing off automated resume screening... Language Processing (SNLP) and it's a big part of resume parsing.

How to Get the Applicant Tracking System to Pick Your Resume
https://biginterview.com/blog/2015/03/applicant-tracking-system.html
The entire process of reviewing and evaluating resumes is evolving fast. ... resume screening software, resume robots, soulless automated resume rejecters, etc.
GDPR Art 22 (1)

Automated Decisions?

Yes

Human Involvement?

No

Legal / Similar significant effects?

Yes

Profiling?

No

Yes

No

No

General data protection principles apply

Art 22 applies
Requirements for General Profiling

1. Lawful, fair and transparent
2. Further processing and purpose limitation
3. Data minimisation
4. Accuracy
5. Storage limitation
6. Data Subject rights
1. Lawful, fair and transparent

- concise, transparent, intelligible and easily accessible information
- Mention (among others):
  - How long will data will be kept
  - With whom will it be shared?
  - User rights
  - Purposes
  - Types of personal data being collected
1. Lawful, fair and transparent

- Fair. Do not:
  - deny people access to employment opportunities, credit or insurance
  - target them with excessively risky or costly financial products.
  - categorize them into profiles that will bring disadvantages
2. Further Processing and Purpose Limitation

Further use of data is only allowed if purposes are compatible.
3. Data Minimization

• Aggregate / Pseudonymize where possible
  • Identifiable: Name= John Doe, Salary= 2500EUR, Address= Blaak 45, Rotterdam
  • Pseudonymized: Name = XXX123, Salary=2500EUR, City=Rotterdam
  • Aggregated: AverageSalary=2500EUR, City=Rotterdam

• Minimize volume / # attributes:
  • Minimize storage
  • Minimize use/purpose
  • Minimize access, # recipients
  • Minimize visibility
4. Data Accuracy

• Check accuracy of the data during:
  • collecting data
  • analysing data
  • building a profile for an individual; or applying a profile to make a decision affecting the individual.

• What are possible sources of error?

• Allow the individuals to correct their data easily

“It’s amazing how we’ve transformed the industry as a result of my typo.”
5. Storage Limitation

• Minimize storage:
  • Erase data/parts of data when no longer needed/useful for the purpose
6. Data Subjects Rights

• Right to be informed:
  • Existence of profiling and a decision being made based on profiling

• Right of access:
  • Profile (output), input data, which segments the user is categorized into

• Right to rectify/erase/restrict:
  • Allow individuals to challenge the accuracy of the profiles and the input for the profile
  • Allow individuals to ask for erasure
  • Allow individuals to ask for restriction (temporary freeze)

• Right to object:
  • Allow individuals to object to profiling.
  • Stop profiling unless there is a compelling reason
Requirements for Automated-Decision Making without human involvement

1. Forbidden, unless there are exceptions
2. Transparency
3. Appropriate safeguards
1. By Default Forbidden

- Exceptions:
  - Local law requires it
    - Preventing fraud and tax evasion
  - Necessary for a contract
    - High level of applications (tens of thousands) for a job
  - Explicit consent
- For sensitive data the following exceptions:
  - Explicit consent
  - Necessary for substantial public interest
2. Transparency

• Inform about automated decision making

• Provide meaningful information about the logic
  • Not full disclosure of the algorithm
  • Individual should be able to understand the reason for the decision
  • Case: loan application

• Explain the significance and envisaged consequences of the processing
  • Case: Insurance tracking driving behavior
3. Appropriate Safeguards

• Allow the individuals to:
  • Obtain human intervention
  • Express their point of view
  • Contest the decision
    • Responsibility in case of harm
    • Decision on changing the algorithm

• Frequent assessments on the data sets for any bias
3. Appropriate Safeguards

- Develop ways to address any prejudicial elements, including any over-reliance on correlations
  - Realistic worst case scenarios
  - Who are the users and vulnerable groups?
- Algorithmic auditing, reviews of accuracy/relevance of automatic decision making
More Information

- Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (wp251rev.01)
- General Data Protection Regulation
- Fairness, Accountability and Transparency in Machine Learning