Ethical Machine Learning

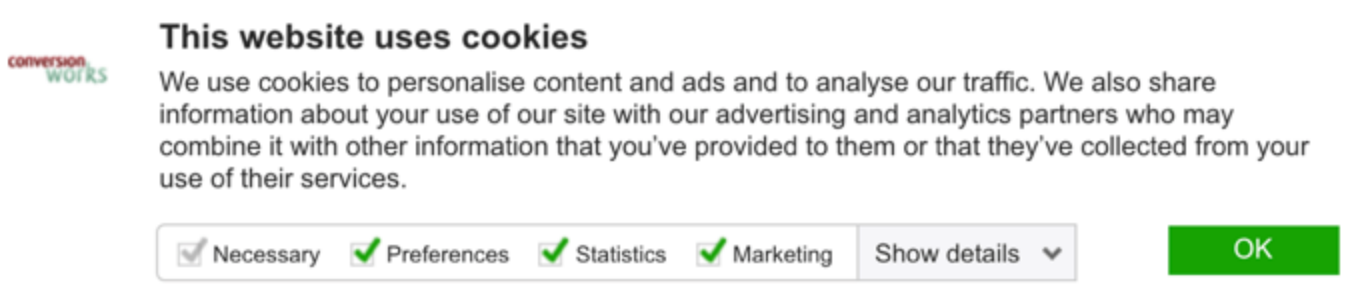
7th March 2022

Diana.pfau@aalto.fi

Agenda

- Overview over the GDPR – personal data, processing principles
- Snapshots – what is allowed ?
- Automated decision-making under the GDPR
- Ethical problems arising form the use of Machine learning and AI

What do you know about the GDPR already?



Personal data (Article 4 (1) GDPR)

- personal data’ means any information relating to an identified or identifiable natural person (‘data subject’);

This includes:

- Name, surname
- Home address
- Email (if: name.surname@....)
- IP address
- Cookie ID

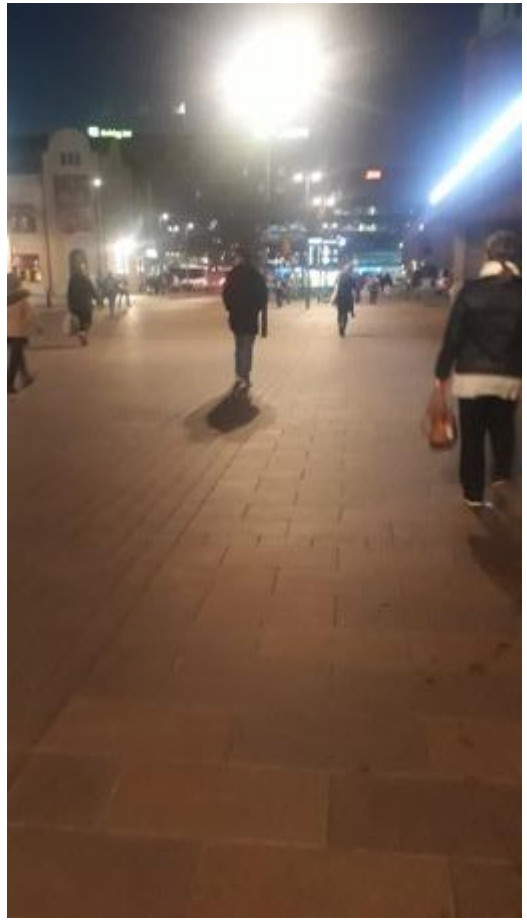
Influences of the GDPR - taking snapshots

- Snapshots can display persons – pictures can include personal data
- Storing snapshots including personal data
- Making public (e.g. uploading pictures on social media) can infringe rights of the persons on the picture
- In general : The GDPR does not apply to the processing of personal data by a natural person in the course of a purely personal or household activity! (Meaning not in connection with a commercial or professional activity)

Personal data on pictures

- Personal data is any information that relates to an identifiable individual. A picture therefore is to be seen as personal data as soon as a person can be identified.
- Consider metadata – timestamp, location,....

Which picture contains personal data?



Household Exemption in the GDPR (Art.2(2))

- Household exemption? In the Netherlands a court found that a grandmother was not allowed to post pictures of her grandchildren on Facebook without being given consent by their mother as their legal representative
- For social networking systems – advancing commercial, political or charitable goals do not fall under the household exemption
- There are reasons to believe that household exemption does not apply
- Be Aware of Constitutional Rights: The right to one´s own picture

Guidance Art. 29 Working Party

- Not all processing on personal data on the internet shall fall outside of the GDPR
- journalistic and artistic expression, social networking blur the lines
- To determine whether internet use falls under the Household exception:
  - Is the personal data disseminated to an indefinite number of persons?
  - Is the personal data about individuals who have no personal or household relationship with the person posting it?
  - Does the scale and frequency of the processing suggest professional or full-time activity?
  - Is there evidence of a number of individuals acting together in a collective and organised manner?
  - Is there the potential adverse impact on individuals, including intrusion into their privacy?

Lindqvist Case

- Mrs Lindqvist was part of a religious group, maintaining a website with personal details of parishioners, greetings to a person that has fallen ill (website established 1998)
- In question: Did Mrs. Lindqvist infringe the Directive 95/46 (household exemption among others?)
- The Court considered:
  - Household exemption cannot be applied in the course of a charitable of religious activity per se

The GDPR in Relation to other rights

- The GDPR as a Regulation dealing with Data Protection finds its justification in the Right to Privacy being a Human Rights and the Right to Data Protection being a Fundamental Right within the EU.
- The GDPR is not the only applicable instrument for the Protection of one´s data and will be supported by several other legislations.
- Germany: The Right to one´s own Picture is a Constitutional Right, therefore the question whether the household exeption would be applicable here, does not matter when we take pictures of German people in Germany. (there is no EU wide reach)

Let´s assume, the household exemption does not apply

- Data Protection Authorities recommed that users of social networking sites shall share pictures about others only if they consented.

- In this case, the person sharing the snapshot online (e.g. on Twitter) is a controller (entity controlling personal data)

A controller has to seek consent for the processing of personal data!

The GDPR applies

- Rules on the collection and use of personal data and special categories of data.
- The processing of personal data follows the following principles:

  - Data Minimization
  - Storage Limitation
  - Lawfulness, fairness, and transparency
  - Purpose limitation
  - Accuracy
  - integrity and confidentiality
  - Accountability

If you used the snapshots for ML

- Scientific exemption?
  - Member states decide upon this exemption in line with the GDPR
  - In this case, right to recitifcation, right to access, right to restriction of processing, and to object can be limited
  - Appropriate safeguards shall be applied, where possible data should be **anonymised** or pseudonymised

- If commercial purpose...

Use of snapshots for ML

In any case, there needs to be a legal basis for the use of snapshots in ML when personal data is involved.

Legal bases (Art. 5 GDPR)

- **Consent**
- **Performance of a contract**
- compliance with legal obligation
- protection of vital interests
- for the perfomance of a task carried out in the public interest or exercise of official auhtority
- **legitimate interest**

Automated decision-making and profiling

The GDPR names several special requirements for:

- automated individual decision-making (= making a decision solely by automated means without any human involvement) and
- Profiling (=automated processing of personal data to evaluate certain things about an individual)

Article 21 GDPR

- The data subject shall have the right to object, on grounds relating to his or her particular situation, at any time to processing of personal data concerning him or her which is based on necessity for tasks carried out in public interest or based on legitimate interest
- The controller shall then no longer process personal data unless it demonstrates legitimate interests outweighing the interests of the data subject
- Data subjects may object to processing at any time when it comes to direct marketing purposes

Article 22 GDPR

- The data subject shall have **the right not to be subjected** to decisions **based solely** on automated processing, including profiling, which **produces legal or similar effects** for him or her.

When would this be the case?

- Mortgage grants ?
- Exclusion from insurances?
- Advertisement?
- Personalization of news?

Structure Article 22

(1) Right not to be subjected to decisions based solely on automated decision-making

(2) Exceptions: Necessary to for entering into or performance of a contract, where member states law requires, or explicit consent was given

(3) Where (2) Is applicable, at least additional safeguards and right to human intervention on the side of controller to give his or her opinion

(4) Special categories of data may not be processed (=health, political/religous believes, tradeunion membership, sexual orientation)

Article 22 GDPR

- **The right not to be subject to solely automated processing**

- **“to meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject”**

Edwards, Lilian and Veale, Michael, Enslaving the Algorithm: From a ‘Right to an Explanation’ to a ‘Right to Better Decisions’? (2018). IEEE Security & Privacy (2018) 16(3), pp. 46-54, DOI: 10.1109/MSP.2018.2701152, Available at SSRN: https://ssrn.com/abstract=3052831 or http://dx.doi.org/10.2139/ssrn.3052831

- **The right to human intervention**

Canellopoulou-Bottis, Maria and Panagopoulou, Fereniki and Michailaki, Anastasia and Nikita, Maria, The Right to Human Intervention: Law, Ethics and Artificial Intelligence (July 31, 2019). Available at SSRN: https://ssrn.com/abstract=3430075 or http://dx.doi.org/10.2139/ssrn.3430075

The data subject has to be informed at the time of data collection of the following

- The identity and contact details of the controller
- The contact details of the Data protection officer where applicable
- The purposes of processing for which the personal data are intended and the legal basis for such processing
- The recipients or categories of recipients of the personal data if any
- Where personal data is transferred outside the Eu- existence of absence of an adequacy decision by the Commission
- The period for which the personal data will be stored (where not possible the criteria used to determine that period)

And the data subject´s active rights

- The right to access
- The right to withdraw consent
- The right to object to processing
- The right to data portability
- The right to rectification and erasure
- The right to lodge a complaint with supervisory Authority
- The existence of automated decision-making AND where it applies. The right to explanation
- Additionally: where processing is based on explicit consent: the right to obtain human intervention on the part of the controller

Consequences for ML with Snapshots

- If the snapshots contain personal data make sure to:
  - Ask for explicit consent (in written) that is based on exhaustive information
  - Determine the purpose of the processing and the storage time
  - Consider anonymising data wherever possible
  - Include additional safeguards to ensure that data does not get stolen or lost
  - Ensure that the model is explainable
  - Where possible, keep human oversight
  - Ensure security of all information you hold

Societal impacts Machine Learning and AI

- Basic assumption on data subjects: informed, reasonable <-> Machine Learning and AI being in the hand of a few “frightful five” basing their operations on acting outside of the people´s knowledge
- Reinforcement of societal biases?
- Division of wealth?
- Access rights?
- Clearview AI – an idea

https://iapp.org/news/a/clearview-ai-plans-to-have-100b-facial-images-in-database-by-next-year/

Reaction EU – draft AI regulation

- Prohibited practices
- placing on the market/use of AI systems using subliminal techniques beyond a person´s consciousness to materially distort a person´s behaviour
- AI systems exploiting vulnerabilities of a specific group of persons due to age, physical or mental disability to distort behaviour of a person which is likely to cause harm to that person
- Use of AI by public authorities to classify trustworthiness of persons or characteristics with the social score leading to either detrimental or unfavourable treatment
- [....]
- https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF

Further Research

- Is focusing on many aspects when it comes to ML :
  - Data Protection by design – how to ensure that model can be compliant with the GDPR
  - Machine Learning and the right to be forgotten
  - Ethical AI – avoidance of discrimination, acknowledging privacy, non-manipulation
  - Interesting pending case: https://www.classlawgroup.com/clearview-data-breach-lawsuit/

Thank you !

Consent

Must be:

- Freely given
- Specific
- Informed
- Indication of choice

Can an employee give consent to CCTV in the office of his employer?

Be aware: made public, combination of data!