Privacy Preserving Machine Learning
Seminar kick-off

April 29, 2021

seminar@m-chair.de

Chair of Mobile Business & Multilateral Security

Goethe University Frankfurt

Agenda

1. Organizational information
2. Introduction to Privacy and Data Protection
3. Presentation of topics
4. Questions

Chair of Business Administration, especially Business Informatics, Mobile Business and Multilateral Security

Chair of Mobile Business & Multilateral Security

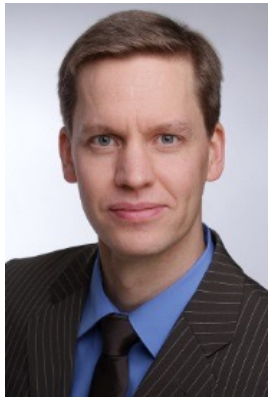
Theodor-W.-Adorno-Platz 4
Campus Westend
RuW, 2nd Floor

Phone: +49 69 798 34701
Fax: +49 69 798 35004
eMail: info@m-chair.de

www.m-chair.de

Team

Kai Rannenberg



Sebastian Pape

Narges Arastouei

Welderufael Tesfay

Frédéric Tronnier

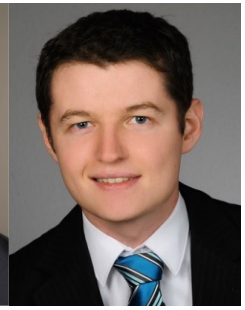
Ahad Niknia

Sascha Löbner

Ann-Kristin Lieberknecht

Christopher Schmitz

David Harborth



Peter Hamm

Supervisor

Sascha Löbner, M.Sc.

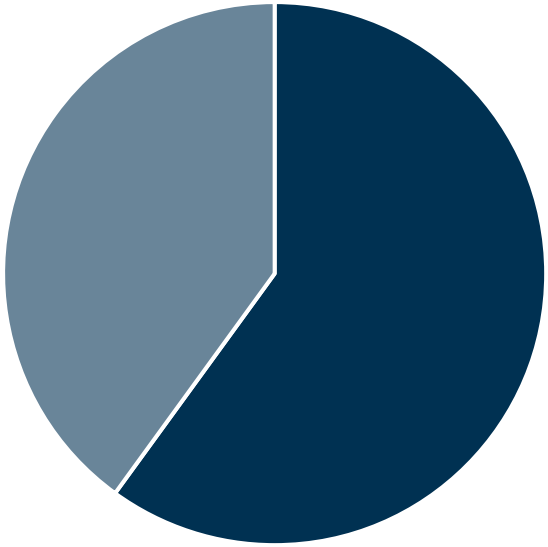
RuW Building, Office 2.236

Email: sascha.loebner@m-chair.de

seminar@m-chair.de

Seminar grade

- This seminar consists of two administrative parts:



- 60% Paper
- 40% Presentation

- Participation in all parts is **required** for the successful completion of the seminar.
- The work is evaluated on an individual basis.

Formal requirements

- For the paper, the formal requirements of the chair apply.
  - Please use the provided word template (or LaTeX)
  - Use the APA American Psychology Association style for citations
  - 10 pages text are recommended (excluding cover, table of contents, references, etc.)

Submission

- The seminar papers must be submitted in **electronic form** in the following format:
  - Ms-word/OpenOffive/LaTeX.zip AND
  - Adobe PDF (Make sure that the file can be opened with Adobe PDF Reader)

  via E-Mail to: seminar@m-chair.de

- The PDF fiile should include the statutory declaration with **your scanned signature**

- Submission until 4th June 2021

Formal requirements for presentations

- Seminar presentation:
  - Duration: 15 min. at most
  - Following discussion: 15 min
- Each presentation is assigned a moderator
  - Responsible for the first question
  - Moderating the discussion
- Submission until 9th June 2021
  - PDF or PPTX
  - Email to seminar@m-chair.de

Important dates

Date	What	Where/When/How
29th May 2021	Kick-off	Big Blue Button
30th May 2021 (12:00 pm)	Submission of preferred topics (1-3)	Email to: seminar@m-chair.de
30th May 2021	Distribution of topics	Via email
4th June 2021 (by midnight)	Paper submission	Ms-word / OpenOffive / LaTeX.zip AND PDF to: seminar@m-chair.de
9th June 2021 (by midnight)	Presentation submission	Email to: seminar@m-chair.de
10th June 2021	Presentation – day 1	09:00 - 17:30
11th June 2021	Presentation – day 2	09:00 - 17:30

Possibly one of the presentation days will be cancelled or shortened!

Contact

In case of any questions or problems arise during the seminar you can contact: seminar@m-chair.de

For comprehensive questions please make an appointment for your topic:

- sascha.loebner@m-chair.de

Agenda

1. Organizational information
2. Introduction to Privacy and Data Protection
3. Presentation of topics
4. Questions

mobile business

Agenda

1. Organizational information
2. Introduction to Privacy and Data Protection
   - Introduction
   - Legal aspects
   - User aspects
   - Technical aspects
3. Presentation of topics
4. Questions

Data Protection and Privacy

- Both terms are related but not synonymous and have many definitions.
- 2 popular ones:
  - Data protection is the protection from harmful and unsolicited usage of data linked to the personal sphere of a person.
  - Privacy is the right to be left alone, e.g. to be unwatched or anonymous [WaBr1980]

Introduction I

- Early day definitions: "The right to be let alone" Warren and Brandeis, 1890, Harvard Law Review: "The right to privacy" [WaBr1890]
- Beginning of information age: "The claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others." Westin, 1967.

- Westin's index
  - Privacy fundamentalists
  - Privacy pragmatists
  - Privacy unconcerned

Source: https://pixabay.com/es/de-distancia-junction-direcci%C3%B3n-1020088/

Introduction II

- Contemporary: **It is complex.**
  - "The ability of the individual to protect information about [herself]" Goldberg et. al 1997

- Personal information: "Any information relating to an identified or identifiable natural person (data subject); an identifiable person is one who can be identified directly or indirectly "

Source: https://pixabay.com/es/icono-la-cabeza-ver-el-perfil-1247948/

mobile business

Agenda

1. Organizational information
2. Introduction to Privacy and Data Protection
   - Introduction
   - Legal aspects
   - User aspects
   - Technical aspects
3. Presentation of topics
4. Questions

General Data Protection Regulation (GDPR)

- Entered into force on 24 May 2016 and applies since 25 May 2018.

- The European Commission says that the recently approved regulation “puts the citizens back in control of their data, notably through”:
  - **A right to be forgotten** - Users will have the right to demand that data about them be deleted if there are no "legitimate grounds" for it to be kept.
  - **Data security:** Personal data that is “any information relating to an identified or identifiable natural person” (GDPR article 4) has to be protected against loss, damage and unauthorized processing

[GDPR 2016]

General Data Protection Regulation (GDPR)

Law Alone is not Sufficient

- Data protection / Privacy law alone not sufficient
  - Not all processing can be controlled (e.g. every network node).
  - Deliberate breaking and bending of law (different legislations on the internet)
  - Economic pressure can force customers to give consent to almost any kind of ‘privacy’ policy (e.g. selling privacy for “peanuts”).

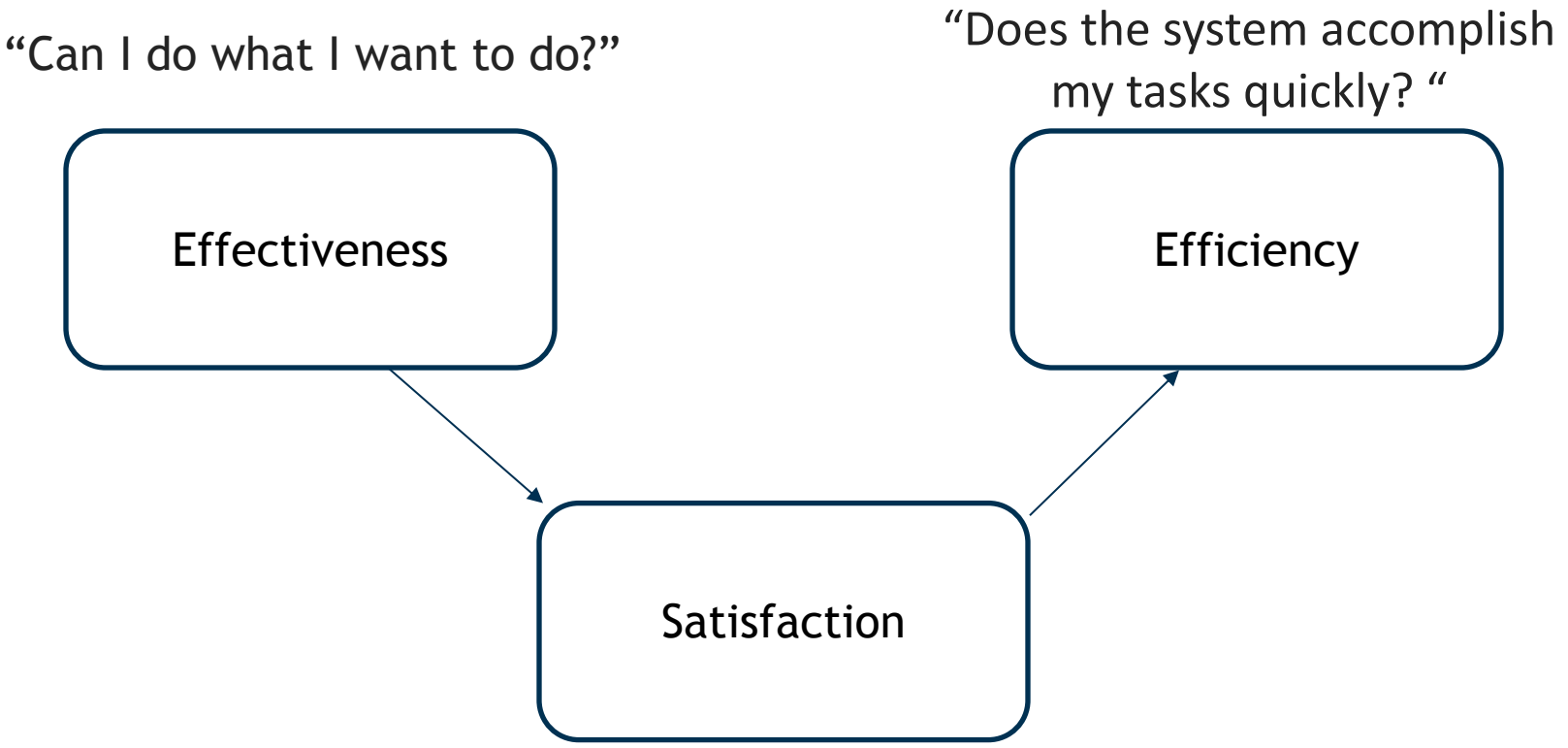
[Reagle1998, SelfReg1999, Bell2001, Hoofnagle2005]

mobile business

Agenda

1. Organizational information
2. Introduction to Privacy and Data Protection
   - Introduction
   - Legal aspects
   - User aspects
   - Technical aspects
3. Presentation of topics
4. Questions

Usability



[National Academy2010]

Agenda

1. Organizational information
2. Introduction to Privacy and Data Protection
   - Introduction
   - Legal aspects
   - User aspects
   - Technical aspects
3. Presentation of topics
4. Questions

Technical Aspects of Privacy

A. Privacy by design
B. Privacy engineering
C. Privacy enhancing technologies

Source: https://pixabay.com/es/humanos-siluetas-redes-internet-1157116/

A. Privacy by design

- Refers to the notion of embedding privacy directly into the design of ITs and systems
- Adopted as one essential principle in the GDPR.

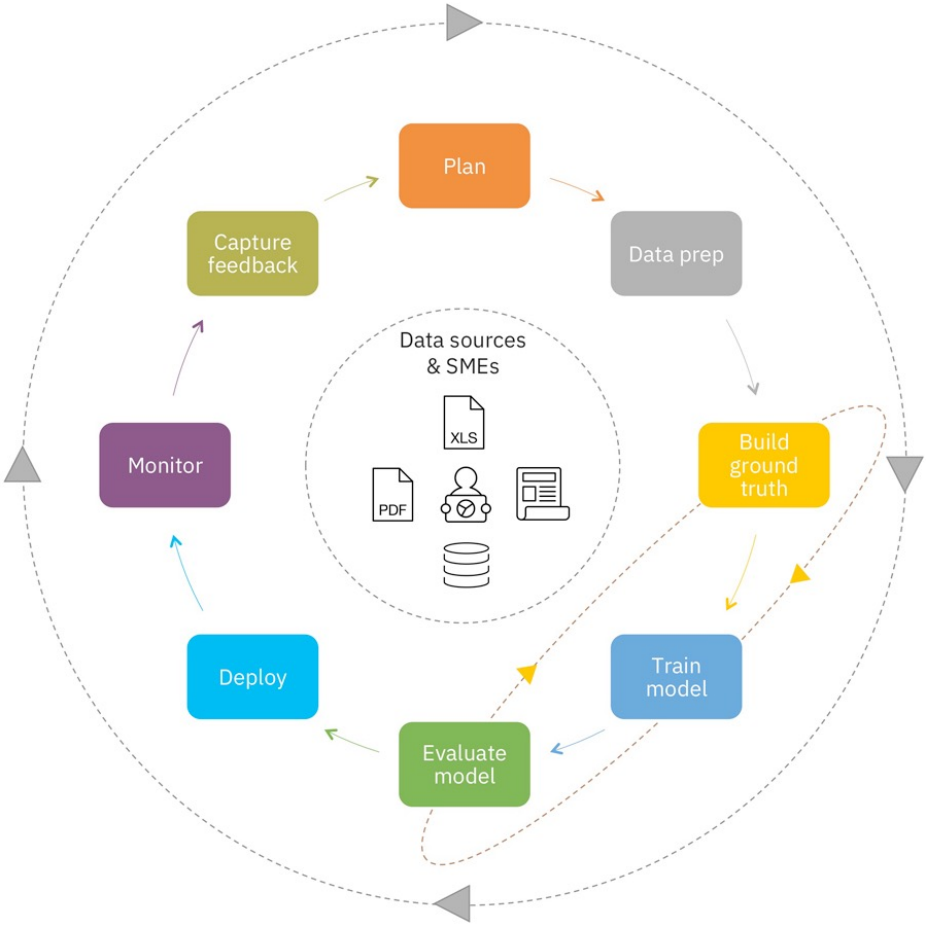
[Cavoukian2010]

7 foundational principles

- Proactive not reactive
- Privacy as the Default setting
- Privacy Embedded into the Design
- Full Functionality
- End-to-End Security
- Visibility and Transparency
- Respect for User Privacy

Machine Learning Life Cycle

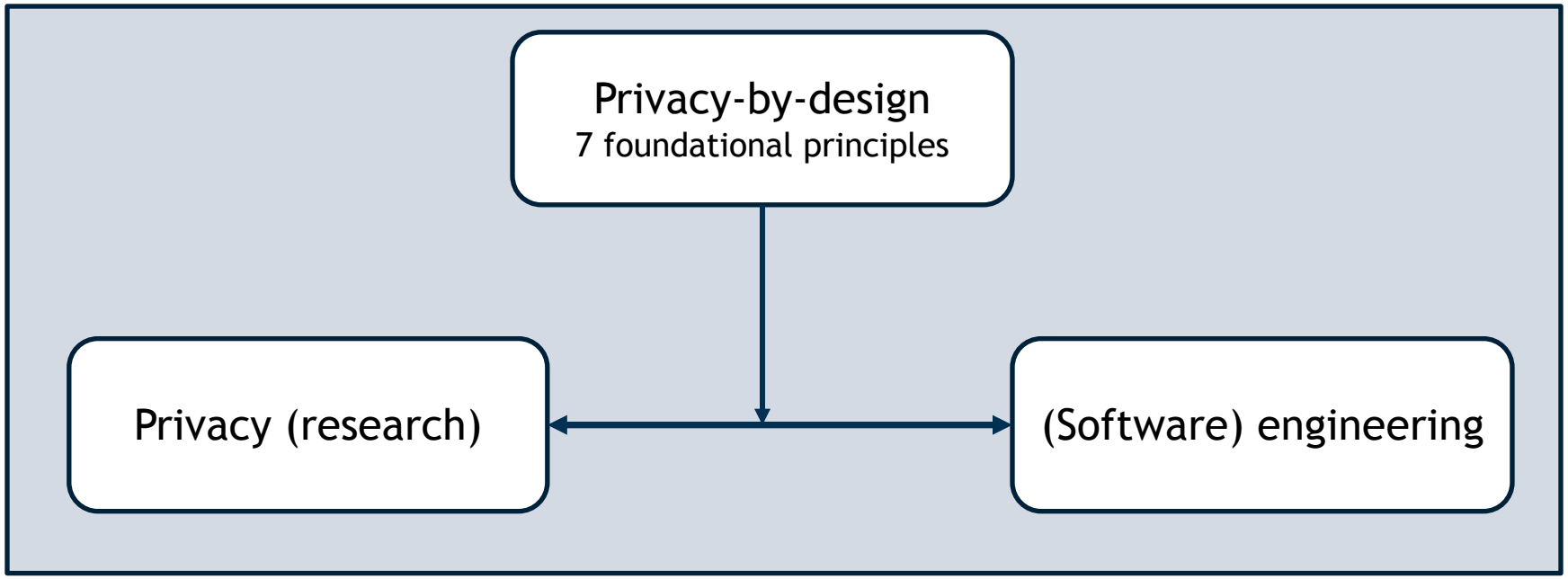
CRISP-DM - additional steps by IBM



Source: https://developer.ibm.com/articles/cc-devops-artificial-intelligence-cognitive/

B. Privacy engineering I

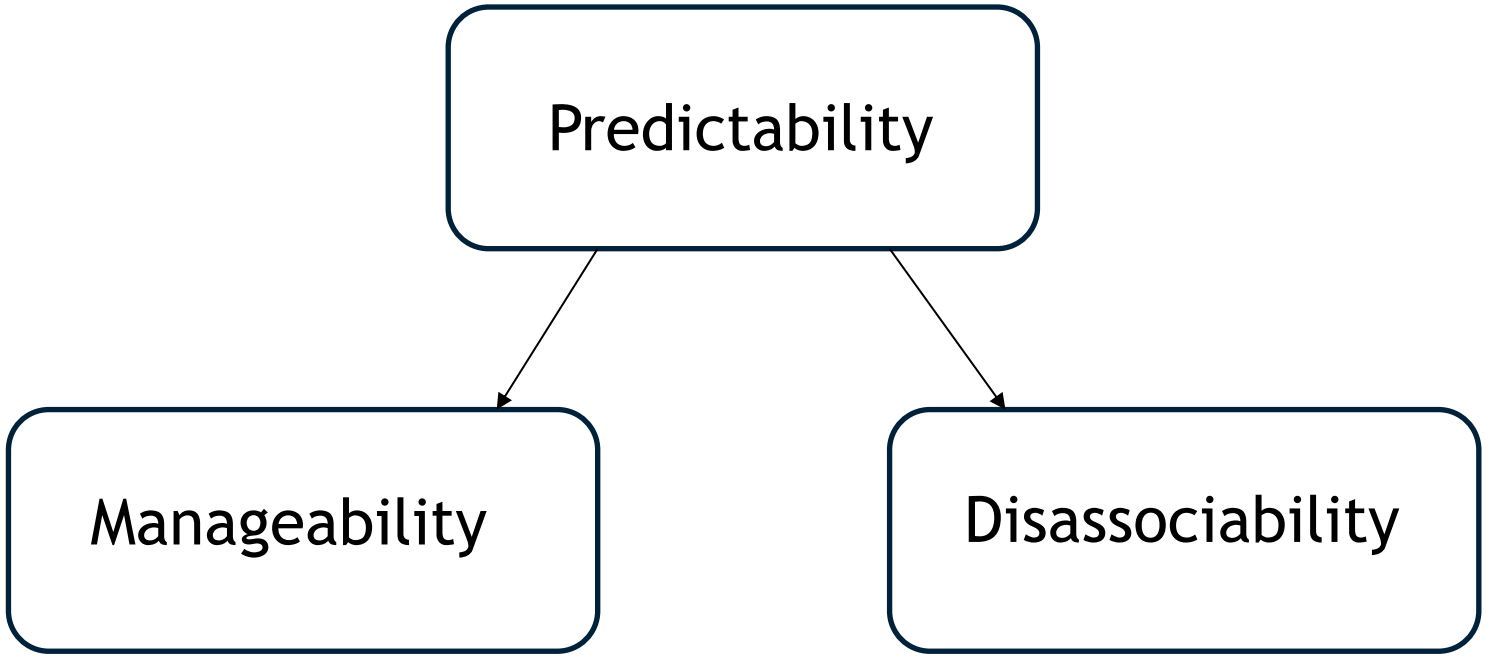
- Connection between research and practice (privacy and software engineering)



[Gürses2016]

B. Privacy engineering II

- Three main goals:



[NIST2014]

C. Privacy enhancing technologies

- Privacy Enhancing Technologies (PETs)
  - It refers to the category of technologies that minimise the processing of personal data
- Examples
  - Automatic anonymisation (e.g. Anonymizer, iPrivacy)
  - Encryption tools (e.g. SSL)
  - Policy Tools (e.g., P3P, TRUSTe)
  - PPML (e.g. Federated Learning, Homomorphic Encryption)

[Danezis2014]

Agenda

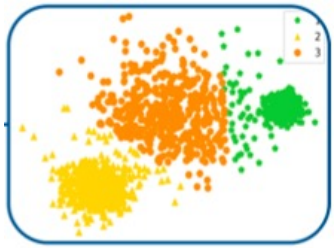
1. Organizational information
2. Introduction to Privacy and Data Protection
   - Introduction
   - Legal aspects
   - User aspects
   - Technical aspects
     - Privacy enhancing technologies for machine learning
3. Presentation of topics
4. Questions

Machine Learning I

- "**Machine learning** is defined as an automated process that extracts patterns from data" [Kelleher2015]
  - **Supervised Learning**: Applications in which the training data comprises examples of the input vectors along with their corresponding target vectors [Bishop2006]
  - **Unsupervised Learning:** The training data consists of a set of input vectors without any corresponding target values.

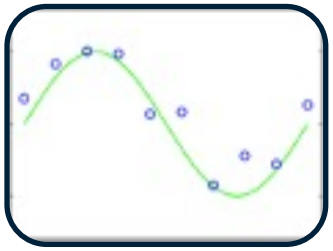
Machine Learning II - Tasks

- **Clustering:** Dividing the dataset into clusters of similar examples. (e.g. spam filter)



- **Classification:** The computer program is asked to specify which of *k* categories some input belongs to. (e.g. object recognition)

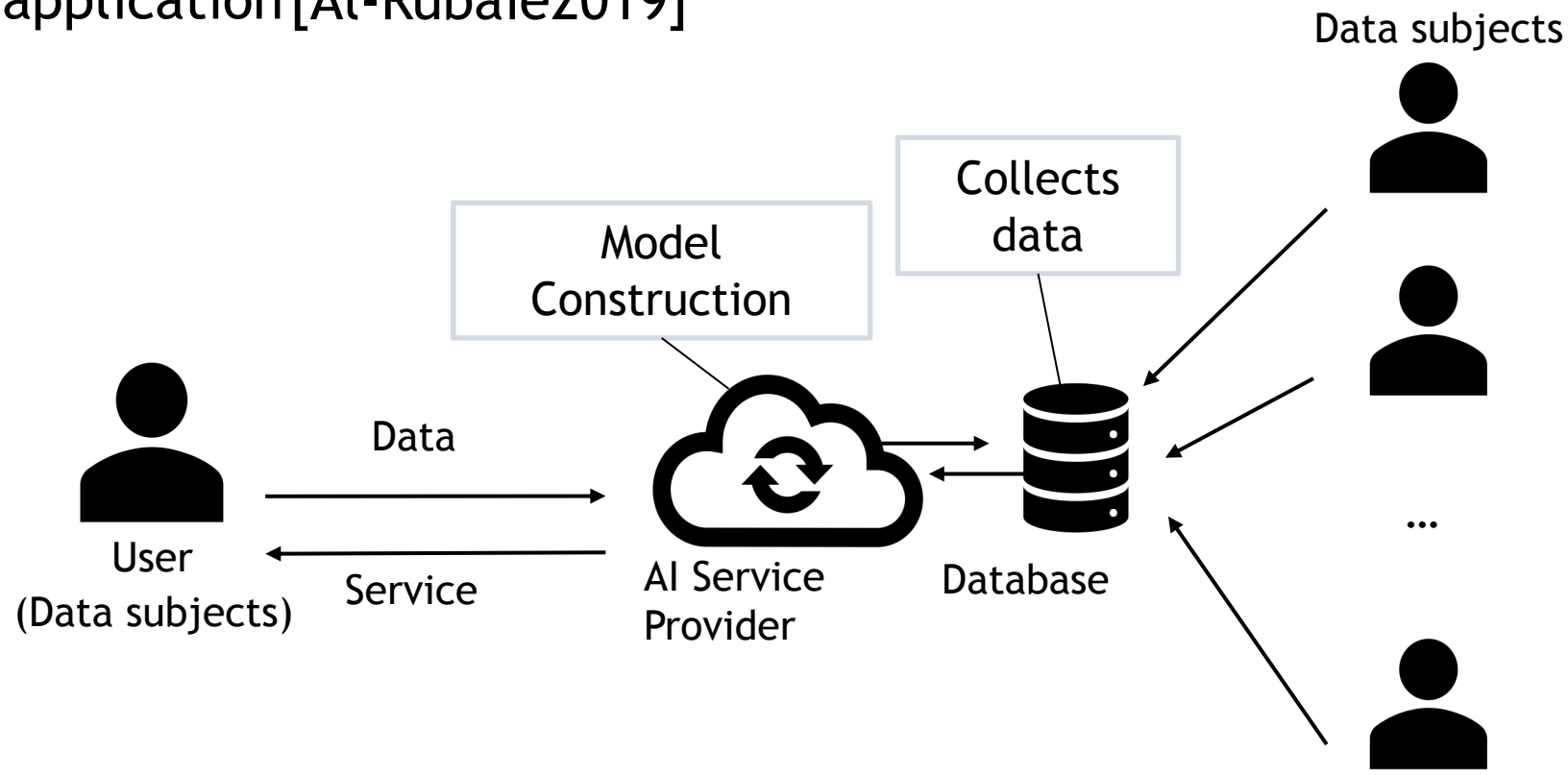
- **Regression:** The computer program is asked to predict a numerical value given some input. (e.g. price prediction)



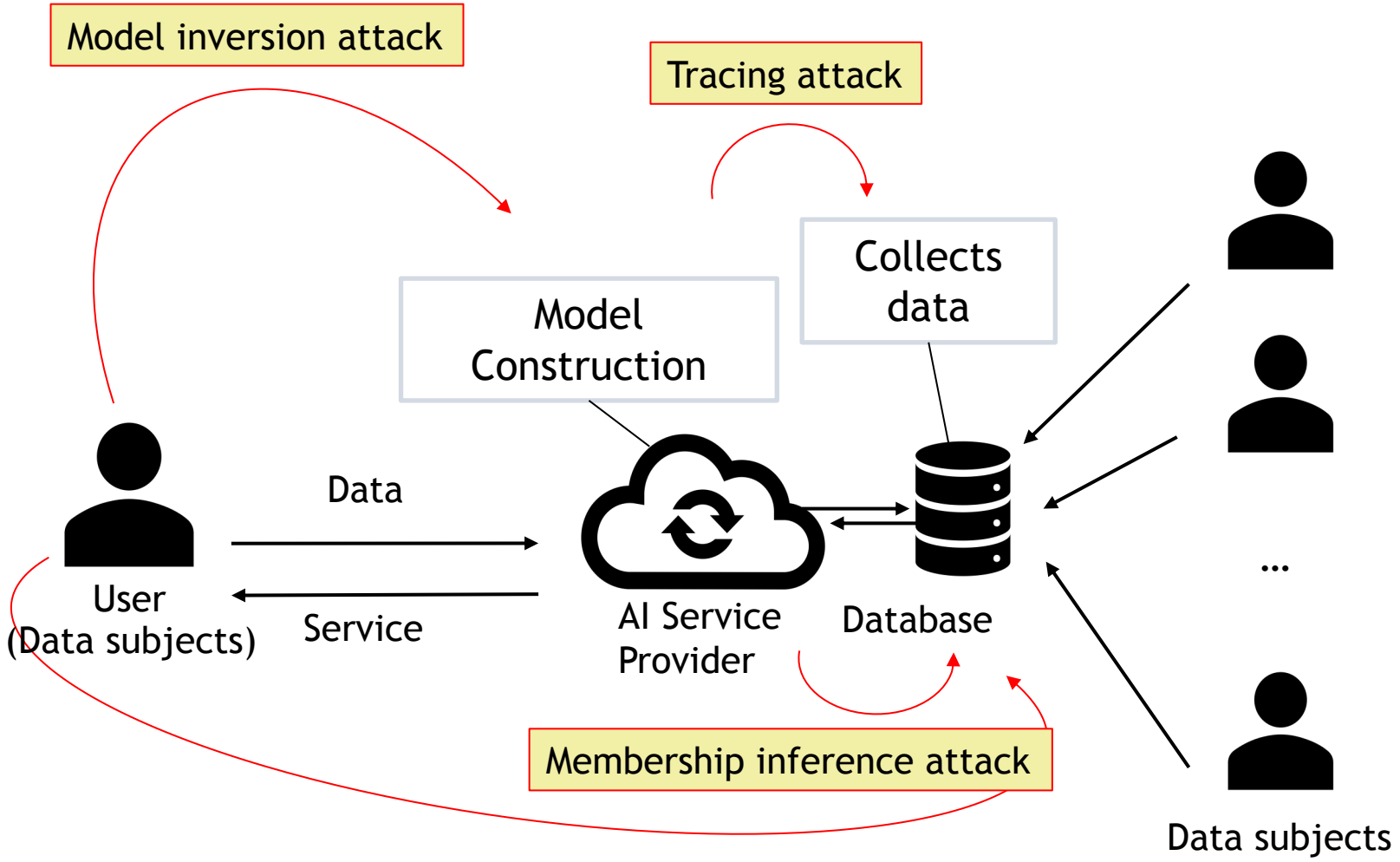
[Goodfellow2016]

Why Privacy Preserving Machine Learning?

- Machine Learning (ML) is becoming part of our daily life
- Often individuals' private data is required for the ML application [Al-Rubaie2019]



Why Privacy Preserving Machine Learning?



--> Law Alone is not Sufficient

Agenda

1. Organizational information
2. Introduction to Privacy and Data Protection
3. **Presentation of topics**
4. Questions

Topic 1: A survey on attacks on machine learning architectures

- Expected results: A review of existing attacks against machine learning architectures and models.
  - What ML architectures exist (e.g. cloud, distributed ML)?
  - What are the trust assumptions in these ML architectures?
  - What are their vulnerabilities?
  - What attacks exist against these architectures?
  - What attacks exist against certain tasks (e.g. clustering, classification)
  - …

Topic 2.1: Privacy preserving federated learning (FL)

- Expected results: A detailed analysis of the technology federated learning.
  - How does federated learning work?
  - How does it increase privacy?
  - Who are the actors involved?
  - What are the trust assumptions?
  - What are the incentives for a user to participate in the training of a local model?
  - What are the strength and weaknesses of federated learning?
  - Where is the technology currently used?
  - ...

Topic 2.2: Privacy preserving differential privacy (DP)

- Expected results: A detailed analysis of the technology differential privacy.
  - How does differential privacy work?
  - How does it increase privacy?
  - Who are the actors involved?
  - What are the trust assumptions?
  - What are the strength and weaknesses of differential privacy?
  - What are the hurdles for implementation?
  - Where is the technology currently used?
  - …

Topic 2.3: Privacy preserving secure multiparty computation (MPC)

- Expected results: A detailed analysis of the technology secure multiparty computation.
  - How does secure multiparty computation work?
  - How does it increase privacy?
  - Who are the actors involved?
  - What are the trust assumptions?
  - What are the strength and weaknesses of differential privacy?
  - What are the hurdles for implementation?
  - Where is the technology currently used?
  - …

Agenda

1. Organizational information
2. Introduction to Privacy and Data Protection
3. Presentation of topics
4. Questions

Questions

Source: Pixabay released under Creative Commons CC0:
https://pixabay.com/es/pregunta-imagen-plaza-556104/

References

- [Cavoukian2010]: Privacy by Design The 7 Foundational Principles Implementation and Mapping of Fair Information Practices, 2010.
- [D' Acquisto2015]: Privacy by design in big data: An overview of privacy enhancing technologies in the era of big data analytics.
- [Gürses2016]: Privacy Engineering: Shaping an Emerging Field of Research and Practice IEEE Security and Privacy, 14:2, pp. 40-46, 2016.
- [NIST2014]: NIST Privacy Engineering Objectives and Risk Model Discussion Draft. Introduction, 2014.
- [Danezis2014]: Privacy and Data Protection by Design – from policy to engineering, 2014.
- [National Academy2010]: Toward Better Usability, Security, and Privacy of Information Technology: Report of a Workshop
- [Europe2006] European Parliament and the Council: Directive 2006/24/EC of the European Parliament and if the council; www.ispai.ie/DR%20as%20published%20OJ%2013-04-06.pdf
- [EC2014] Progress on EU data protection reform now irreversible following European Parliament vote. Accessed at http://europa.eu/rapid/press-release_MEMO-14-186_en.htm on 12.11.2014.
- [EC-Prot-2014] European Commission: Protection of personal data: http://ec.europa.eu/justice/data-protection/index_en.htm
- [WaBr1890] Samuel D. Warren, Louis D. Brandeis: The Right to Privacy, Harvard Law Review; Vol. IV; December 15, 1890, No. 5; http://groups.csail.mit.edu/mac/classes/6.805/articles/privacy/Privacy_brand_warr2.html
- [Bishop2006] Bishop, C. M. (2006). Pattern recognition and machine learning. springer.
- [Goodfellow2016] Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1, No. 2). Cambridge: MIT press.
- [Kelleher2015] Kelleher, J. D., Namee, B. M., & D'Arcy, A. (2015). Machine learning for predictive data analytics. Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies, 1-19.
- [Al-Rubaie2019] Al-Rubaie, M., & Chang, J. M. (2019). Privacy-preserving machine learning: Threats and solutions. IEEE Security & Privacy, 17(2), 49-58.
- [Zheng2019] Zheng, M., Xu, D., Jiang, L., Gu, C., Tan, R., & Cheng, P. (2019, November). Challenges of privacy-preserving machine learning in IoT. In Proceedings of the First International Workshop on Challenges in Artificial Intelligence and Machine Learning for Internet of Things (pp. 1-7).

Chair of Mobile Business & Multilateral Security

Goethe University Frankfurt
E-Mail: seminar@m-chair.de
WWW: www.m-chair.de