



Machine Learning: Privacy, Regulations and Ethical Issues

October 29, 2024

Seminar kick-off

seminar@m-chair.de

Chair of Mobile Business & Multilateral Security

Goethe University Frankfurt

id42024

- 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



<p>E-Finance</p> <p>Prof. Dr. Peter Gomber</p>	<p>Business Informatics (Informatics)</p> <p>Prof. Dr. Mirjam Minor</p>	<p>Business Informatics & Information Management</p> <p>Prof. Dr. Oliver Hinz</p>
<p>Business Ethics & Business Education (associated)</p> <p>Prof. Dr. Gerhard Minnameier</p>	<p>Business Informatics</p> <p>Hon. Prof. Dr. Matthias Zieschang</p>	<p>Economic and Business Education (associated)</p> <p>Prof. Dr. Eveline Wuttke</p>
<p>Business Education</p> <p>Prof. Dr. Helmut Niegemann</p>	<p>Information Systems & Information Management</p> <p>Prof. Dr. Wolfgang König</p>	<p>Business Education</p> <p>Dr. Christin Siegfried</p>
<p>Information Systems Engineering</p> <p>Prof. Dr. Roland Holten</p>	<p>Business Informatics & Microeconomics</p> <p>Prof. Dr. Lukas Wiewiorra</p>	<p>Mobile Business & Multilateral Security</p> <p>Prof. Dr. Kai Rannenber</p>

Team & External PhD Students



Kai
Rannenberg



Narges
Arastouei



Diana
Weiss



Sascha
Löbner



Atiyeh
Sadeghi



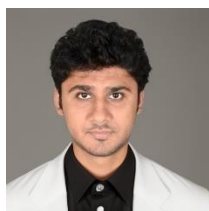
Ann-Kristin
Lieberknecht



Frédéric
Tronnier



Ahad
Niknia



Basharat
Ahmed



Michael
Schmid
Lufthansa



Christopher
Schmitz
Deutsche Börse



Peter
Hamm
Capgemini Invent



Sascha Löbner, M.Sc.

RuW Building, Office 2.236

Email: sascha.loebner@m-chair.de



seminar@m-chair.de

- The **module is passed** if both the **presentation AND** the **term paper** are **passed**.
- The module is **not passed** if the presentation **OR** the term paper is not passed.
 - The **term paper** is passed if it has been assessed with grade of 4,0 or better.
 - **The presentation** has been successfully completed if it has been assessed as "passed".
- The **module grade** results from the **assessment of the term paper**.
- The **presentation** is **NOT included** in the **module grade**.

Formal requirements for writing I

- 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.)

- A good scientific paper should answer the following questions:
 - What is the researched question / problem?
 - What is the concrete new contribution / result of the work?
 - How was this result achieved?
 - Why should the reader believe in it?
- A good scientific paper:
 - addresses a clearly defined question,
 - applies established methods and tools to answer the question,
 - makes this question, the methodology to answer it, the selection decision etc. transparent and thus comprehensible,
 - lives from good documentation and communication

- Scientific papers have standardized structures from which one should only deviate in exceptional cases
 - Abstract may and should already present the results!
 - Introduction
 - Theoretical chapter
 - Background
 - Related Work
 - Methodology
 - Results
 - Discussion
 - Conclusion
 - Textual structural elements should be used as often as reasonable (tables, graphs, "bullet points")
 - Direct citations should be avoided if possible

- **Electronic:**
 - Google Scholar
 - ACM Portal
 - IEEE Xplore
 - Microsoft Libra Academic Search
 - Citeseer Search

- **Bibliothekssuche**
 - UB Frankfurt

- **Further resources:**
 - Ethics Guidelines for Trustworthy AI:
<https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>
 - GDPR: <https://gdpr.eu/tag/gdpr/>
 - AI Act (proposal) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>

Relevant Conferences/Journals

Computer Science Conference:

- IEEE Privacy and Security (A+)
- ACM Conference on Computer and Communications Security (A+)
- ACM Conference on Human Factors in Computing Systems (CHI) (A/A+)
- USENIX Security Symposium (A+)
- Symposium on Usable Privacy and Security (SOUPS) (B)

Information Systems Conference:

- European Conference on Information Systems (ECIS)
- International Conference on Information Systems (ICIS) (A)

Journals:

- MISQ Quarterly
- Computers & Security (COSE)
- Privacy Enhancing Technologies Symposium (PETs)

- The seminar papers must be submitted in **electronic form** in the following format:
 - Ms-word/OpenOffice/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 file should include the statutory declaration with **your scanned signature**
- Submission until 19th of January 2025

- 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 26th of Januar 2025
 - PDF or PPTX
 - Email to seminar@m-chair.de

Important dates

Date	What	Where/When/How
29 th October 2024	Kick-off	RuW 2.202
31 st October 2024 (12:00 pms)	Submission of preferred topics (1-3)	Email to: seminar@m-chair.de
1 st November 2024	Distribution of topics	Via email
19 th January 2025 (by midnight)	Final Paper submission	Ms-word / OpenOffice / LaTeX.zip AND PDF to: seminar@m-chair.de
26 rd January 2025 (by midnight)	Presentation submission	Email to: seminar@m-chair.de
27 th January 2025	Presentation - day 1	09:00 - 18:00
28 th January 2025	Presentation - day 2	09:00 - 18:00
29 th January 2025	Presentation - day 3	09:00 - 18:00
30 th January 2025	Presentation - day 4	09:00 - 18:00
1 st February 2025	Presentation - day 5	09:00 - 18:00

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

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: seminar@m-chair.de

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

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

- 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]

- 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

- 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/>

- 1 Organizational information
- 2 Introduction to Privacy and Data Protection
 - Introduction
 - Legal aspects and guidelines
 - User aspects
 - Technical aspects
 - Ethical issues and fairness
- 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)

THE SIX GDPR PRINCIPLES TO ENSURE ACCOUNTABILITY



- 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”).

- 1 Organizational information
- 2 Introduction to Privacy and Data Protection
 - Introduction
 - Legal aspects and guidelines
 - **User aspects**
 - Technical aspects
 - Ethical issues and fairness
- 3 Presentation of topics
- 4 Questions

“Can I do what I want to do?”

Effectiveness

“Does the system accomplish
my tasks quickly? “

Efficiency

Satisfaction

“Do I feel secure and comfortable
while using the system? “

[National Academy2010]

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

- 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.

7 foundational principles

Proactive not reactive

Privacy as the Default setting

Privacy Embedded into the Design

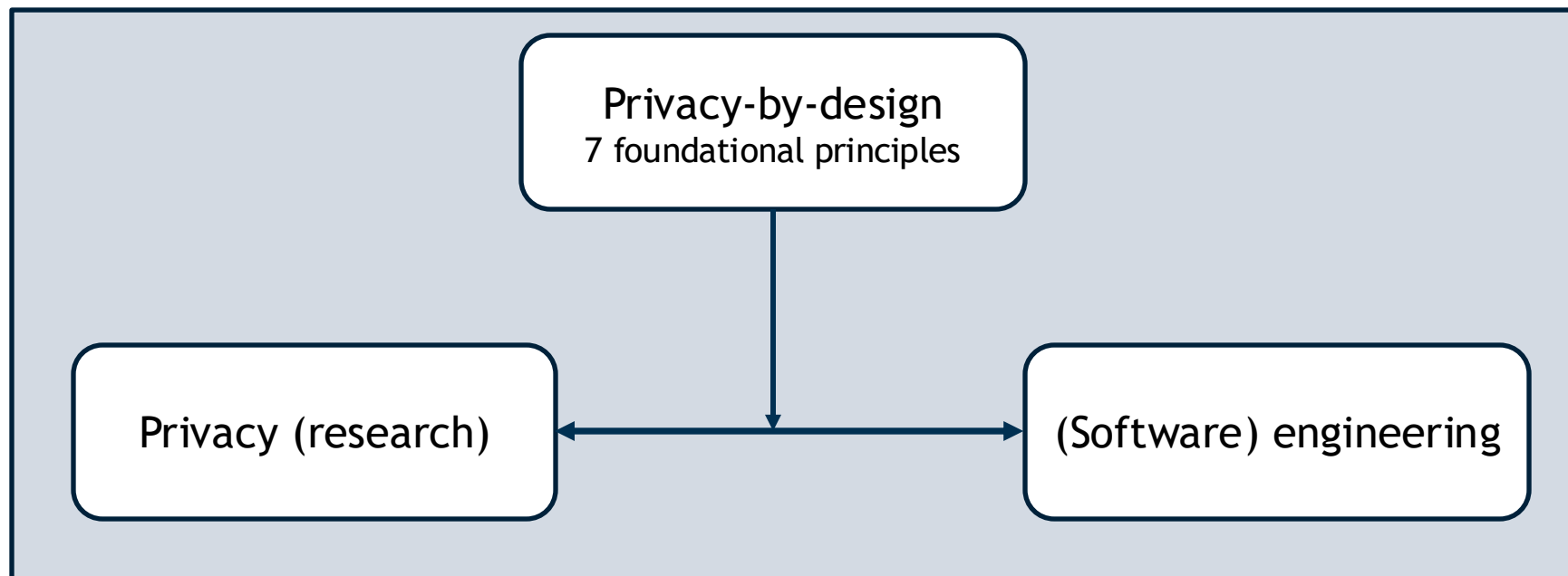
Full Functionality

End-to-End Security

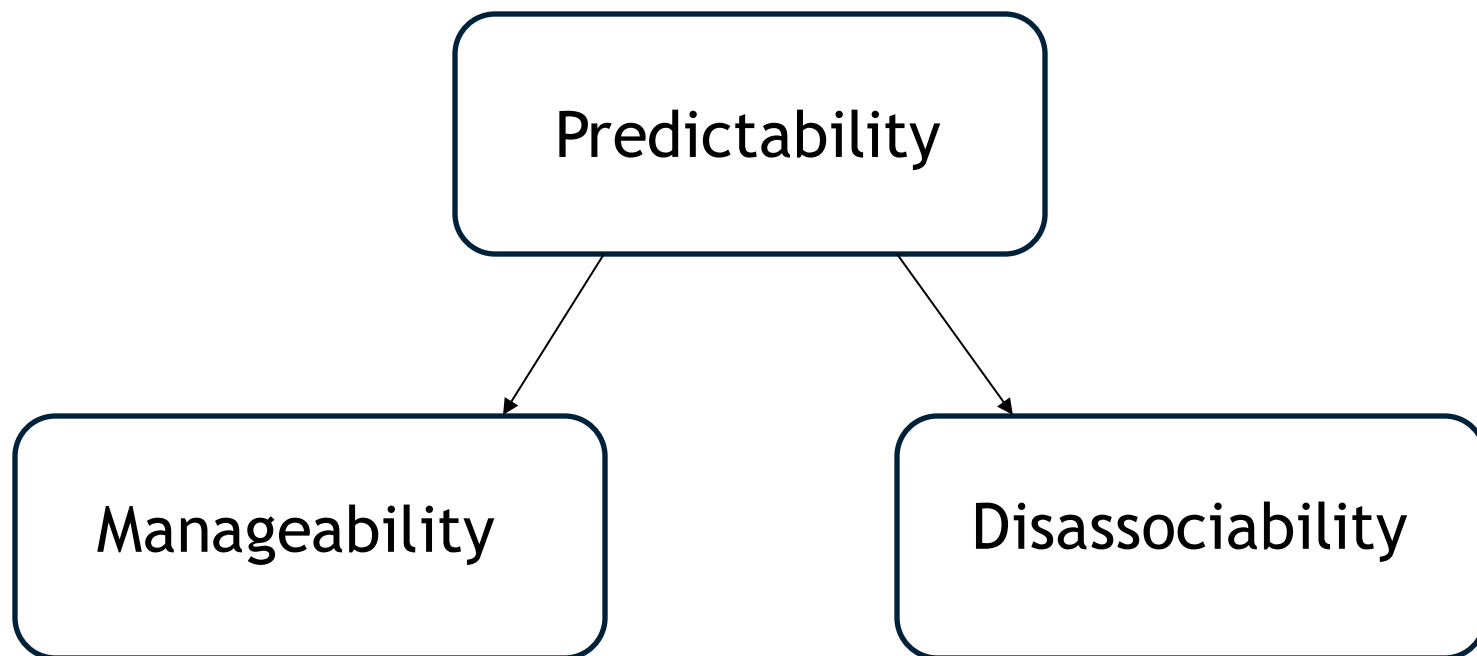
Visibility and Transparency

Respect for User Privacy

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

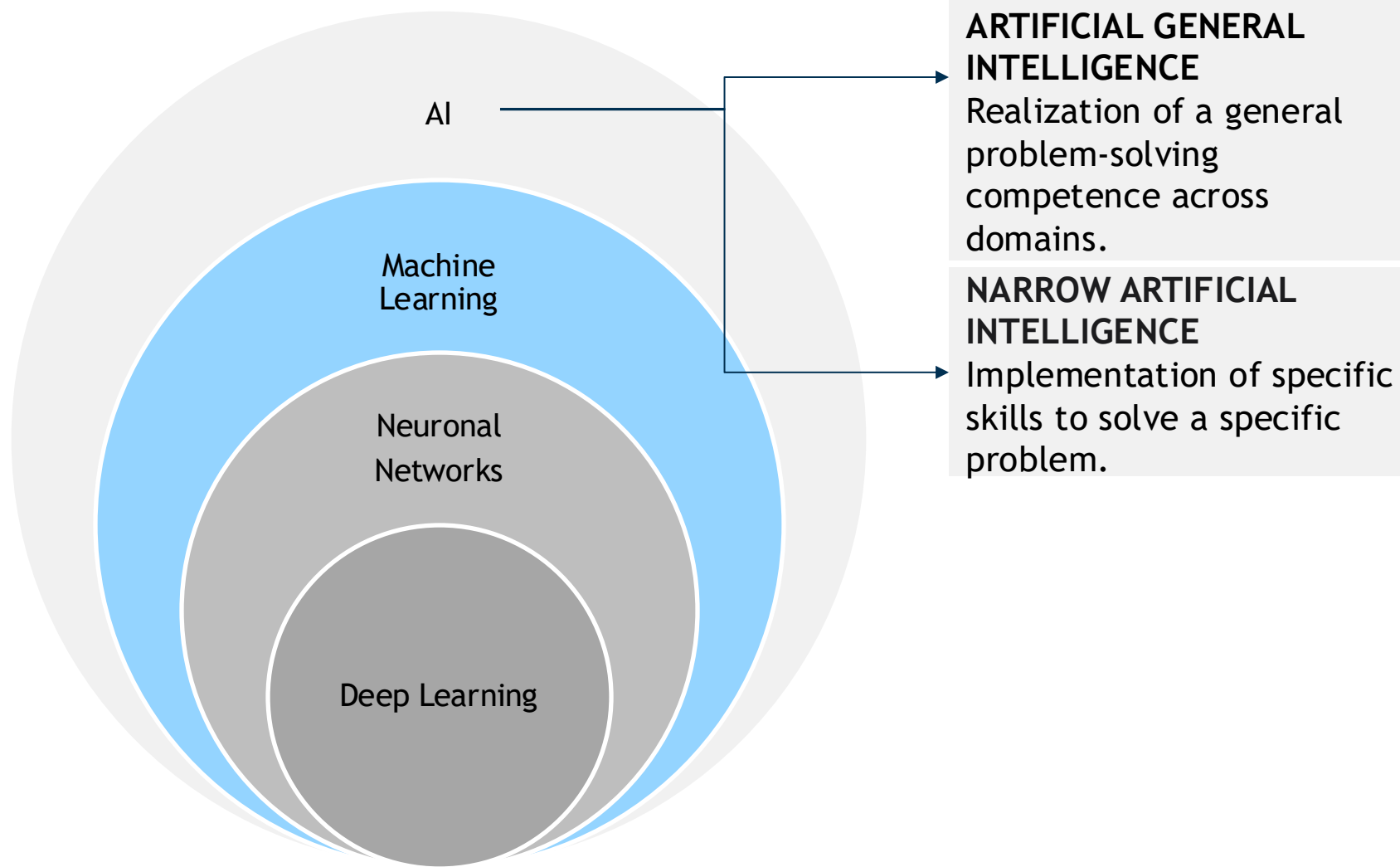


- Three main goals:



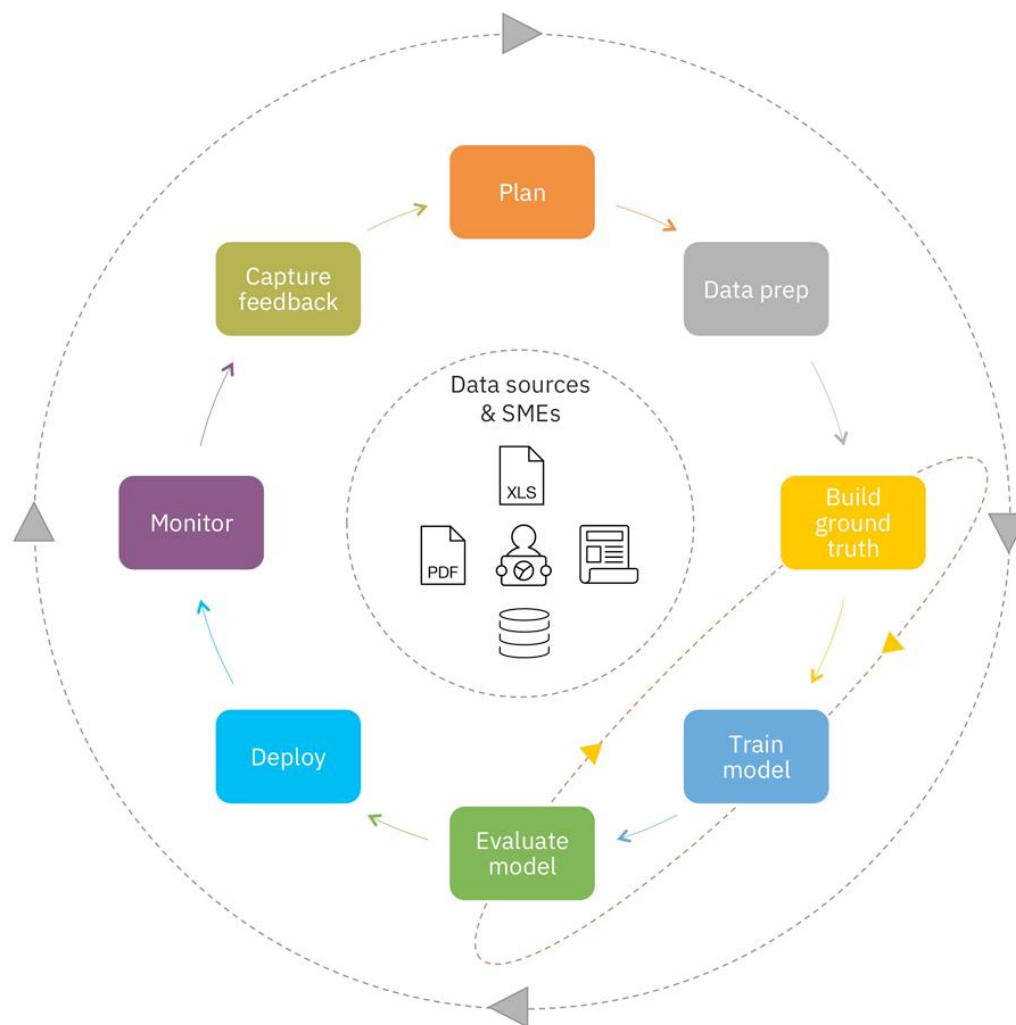
C. Privacy enhancing technologies

- 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. Differential Privacy, Secure Multiparty Computation, Homomorphic Encryption, (Federated Learning))



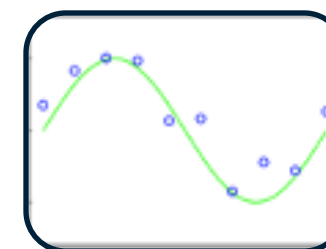
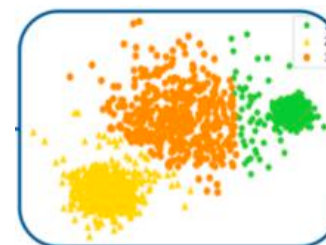
Machine Learning Life Cycle

CRISP-DM - additional steps by IBM



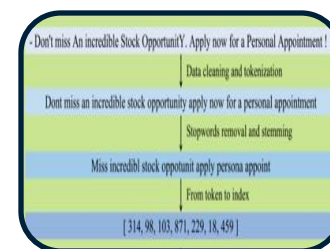
- **“Machine learning is defined as an automated process that extracts patterns from data” [Kelleher2015]**
 - **Supervised Learning:** Applications in which the training data comprises examples of the input vectors along with their corresponding target vectors [Bishop2006].
 - **Unsupervised Learning:**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

- **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)

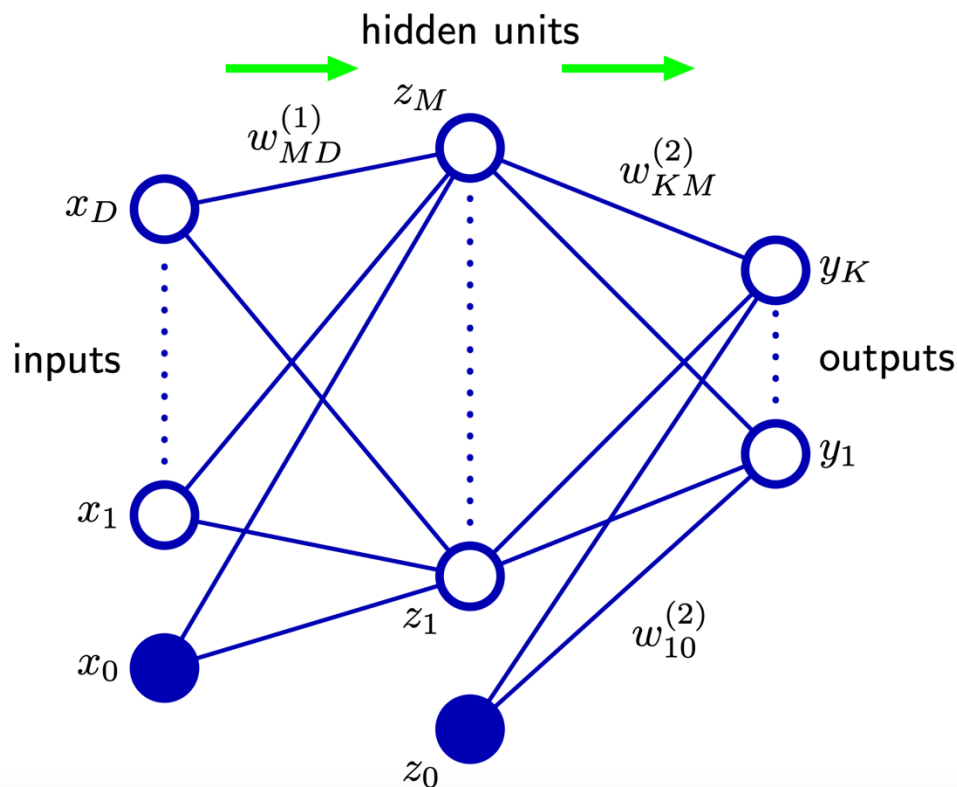


[Goodfellow2016]

- **Natural language processing (NLP):** Use of human languages, such as English or French, by a computer.
- **Speech recognition:** Map an acoustic signal containing a spoken natural language utterance into the corresponding sequence of words intended by the speaker.
- ...



EXCURSUS: Neuronal Networks



$w_{ij} :=$ weights

$w_{i0} :=$ bias

$a_j :=$ activation function

Sigmoid activation function:

$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$

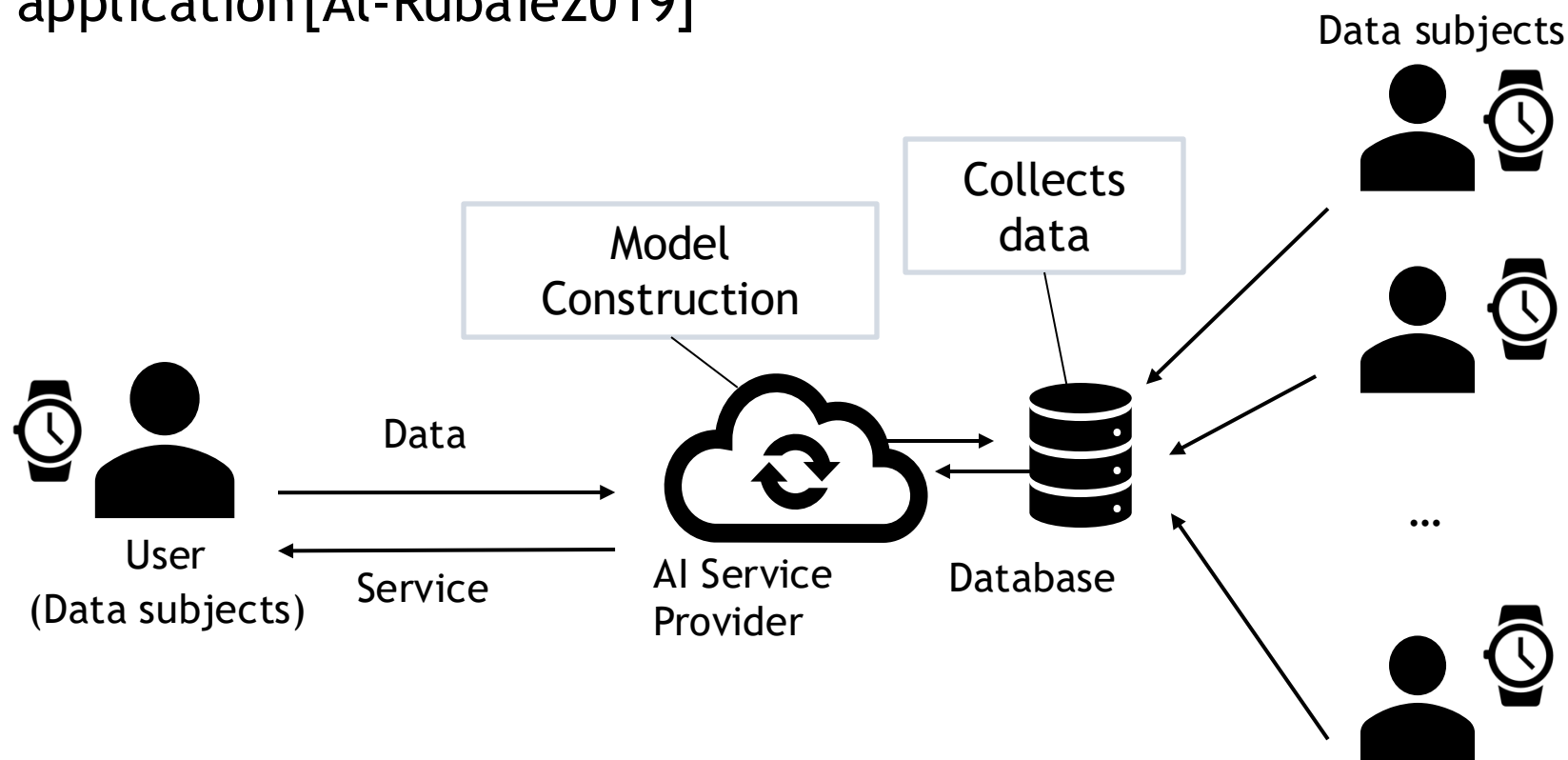
$$y_k(\mathbf{x}, \mathbf{w}) = \sigma \left(\sum_{j=1}^M w_{kj}^{(2)} h \left(\sum_{i=1}^D w_{ji}^{(1)} x_i + w_{j0}^{(1)} \right) + w_{k0}^{(2)} \right)$$

[Bishop06]

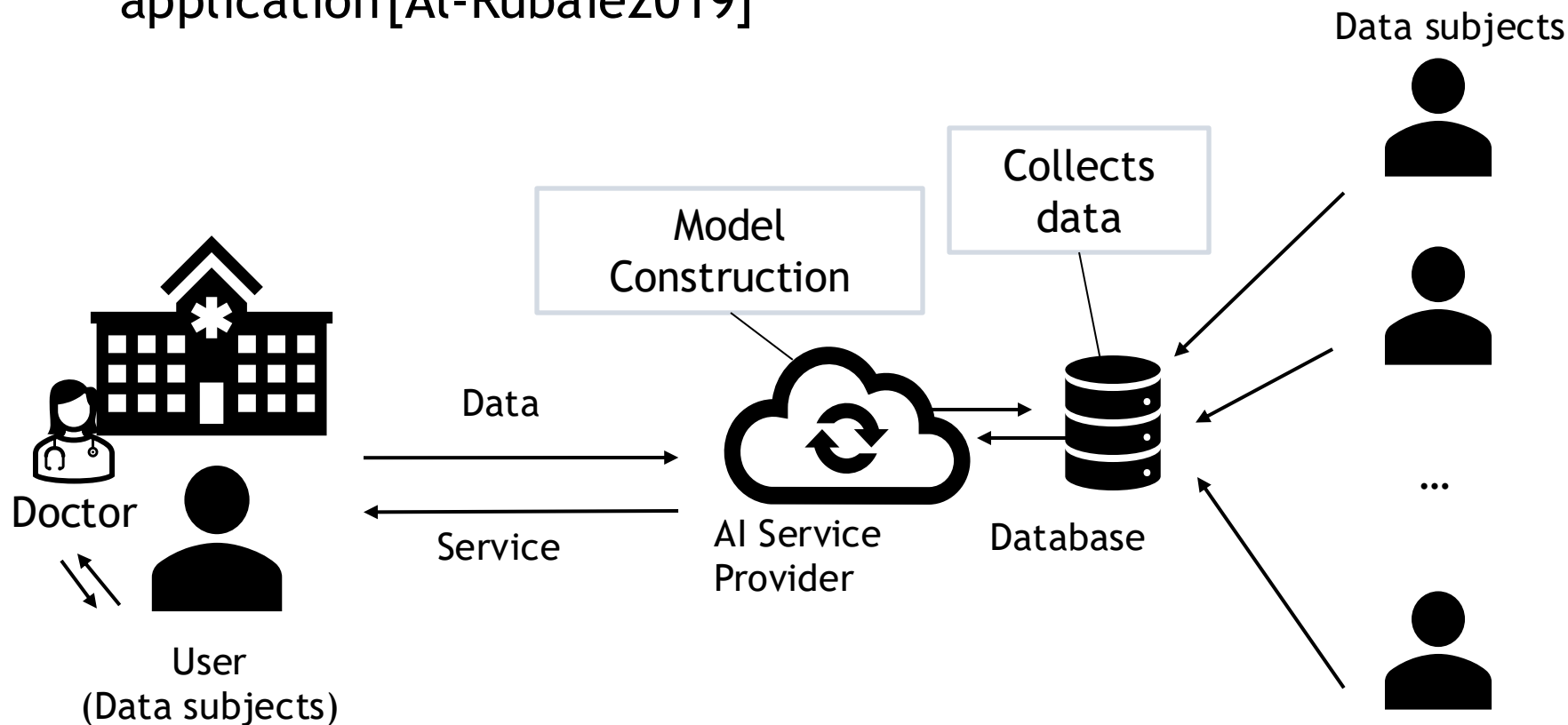
“ChatGPT: Uses a combination of unsupervised pre-training and supervised fine-tuning to generate human-like responses to queries and provide responses to topics that resemble that of a human expert.”

Example Fitness Watch

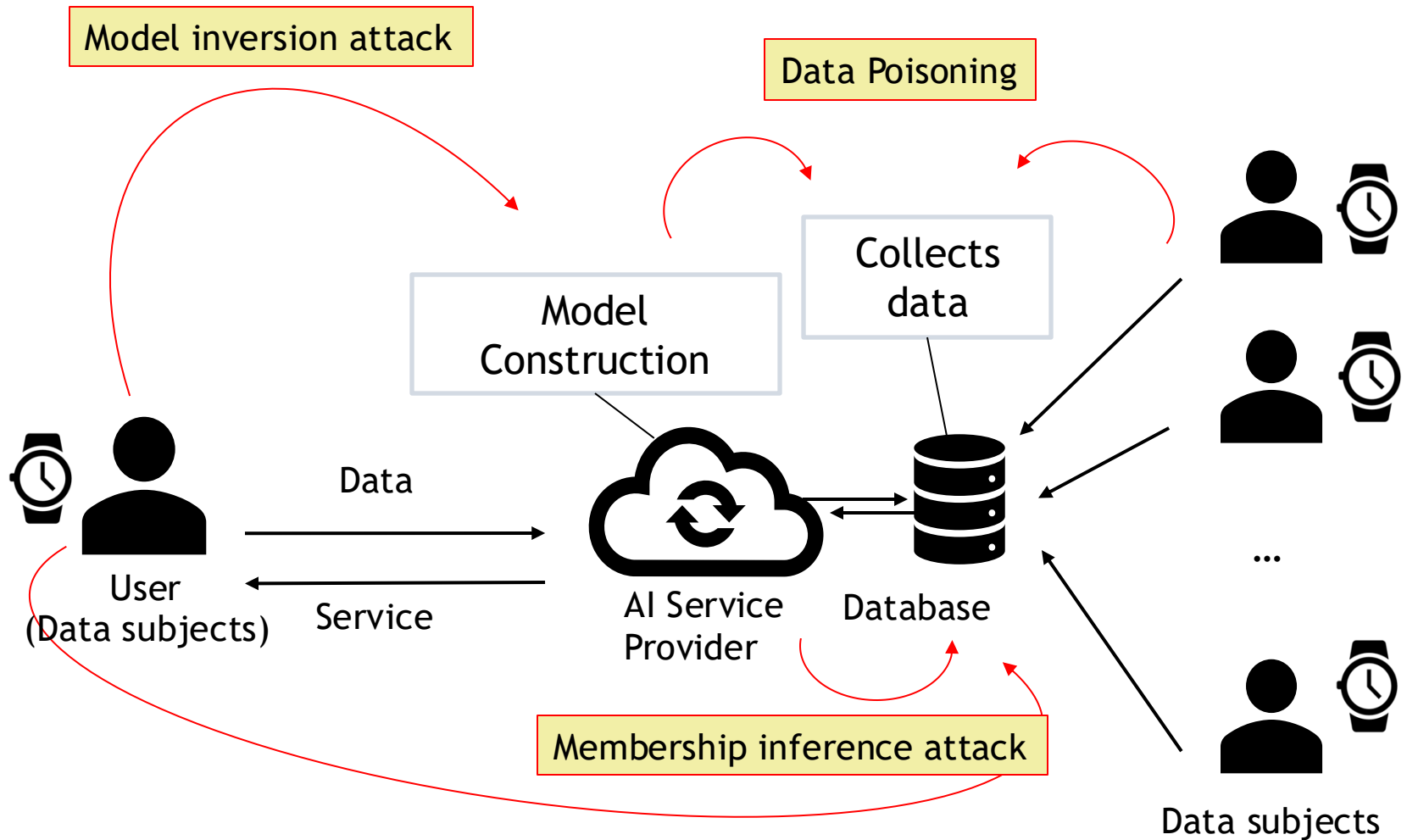
- Machine Learning (ML) is becoming part of our daily life
- Often individuals' private data is required for the ML application [Al-Rubaie2019]



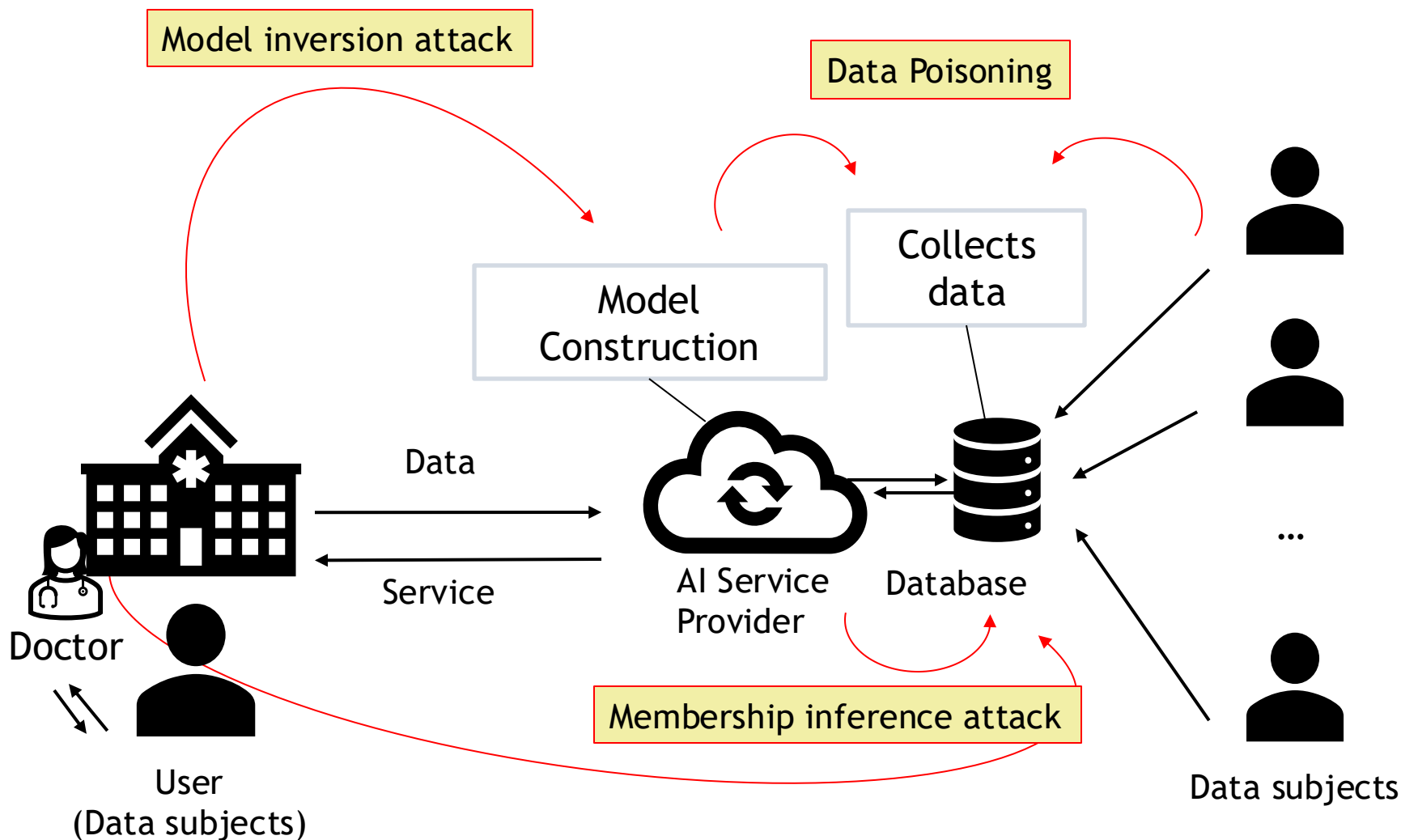
- Machine Learning (ML) is becoming part of our daily life
- Often individuals' private data is required for the ML application [Al-Rubaie2019]



Why Privacy Preserving Machine Learning?



Why Privacy Preserving Machine Learning?



- **Attack:** Poisoning the central model with gradient updates from mislabelled data.
 - Aim 1: Reduce overall accuracy (untargeted) e.g. all diseases
 - Aim 2: Misclassification in a certain class (targeted) e.g. Influenza

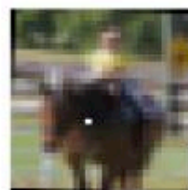
Example: One pixel attack:

- To create adversarial images only the addition of a tiny amount of well-tuned additive perturbation is necessary. Such a modification can cause the image to be labeled into a different class.
- Original class labels are in black
- Modified class labels and the corresponding confidence are blue

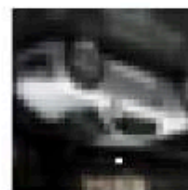
AllConv



SHIP
CAR(99.7%)



HORSE
DOG(70.7%)



CAR
AIRPLANE(82.4%)

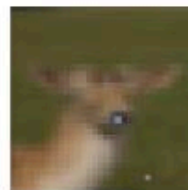
NiN



HORSE
FROG(99.9%)

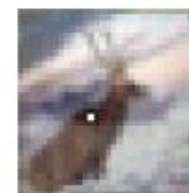


DOG
CAT(75.5%)

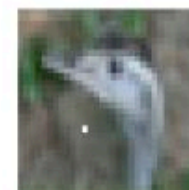


DEER
DOG(86.4%)

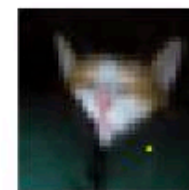
VGG



DEER
AIRPLANE(85.3%)



BIRD
FROG(86.5%)



CAT
BIRD(66.2%)

[Su19]

- Attack: Observing the output to predict training data
- Aim: Determine whether a specific data record was used in the training dataset
- Attacker:
 - Outside
 - Inside e.g. central server or user

Model Inversion Attack



Target



Softmax



MLP



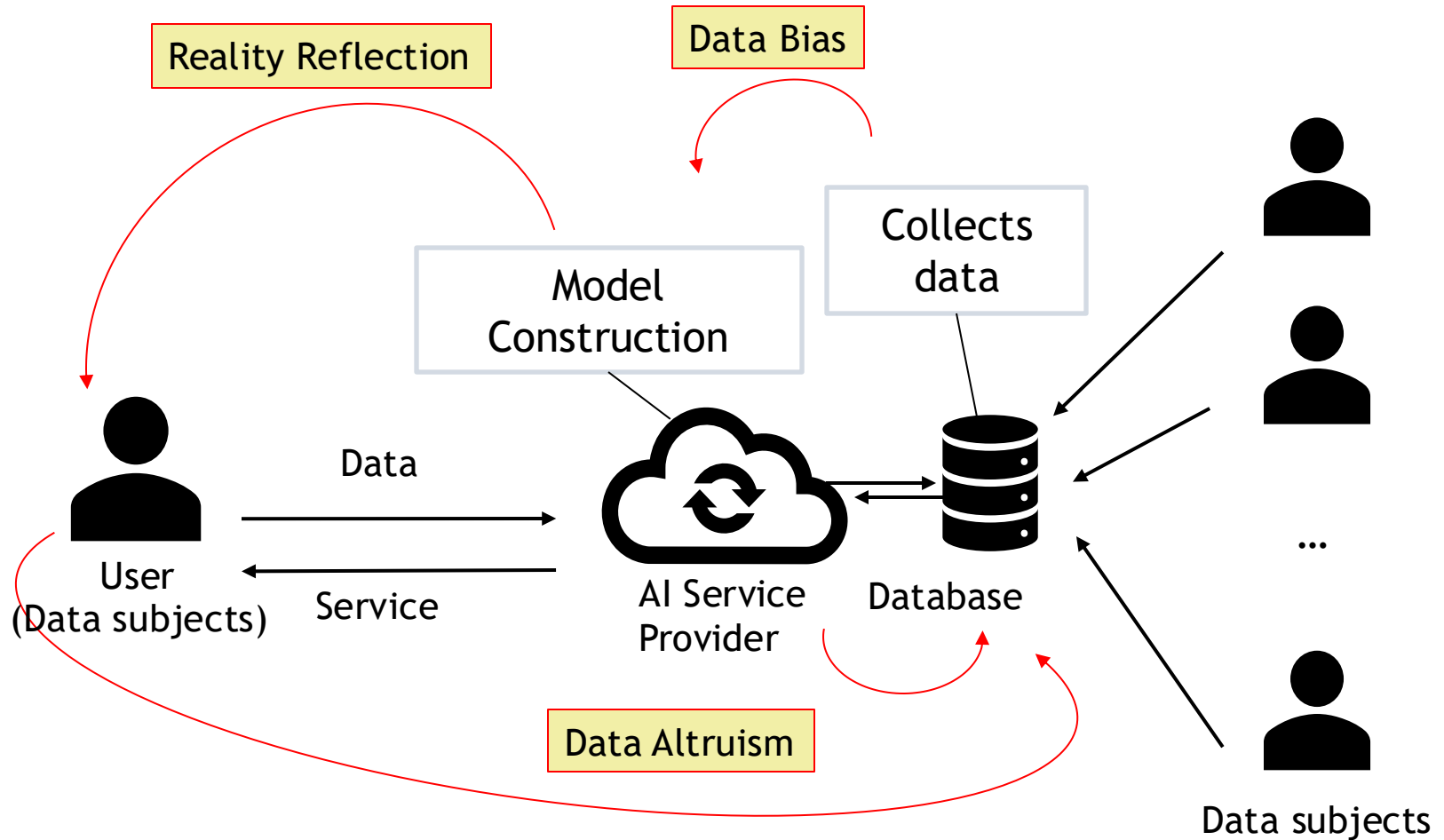
DAE

- Aim to reconstruct private training data from model outputs
- Can extract sensitive information without direct data access
- Require careful balance between privacy protection and model utility
- Defending against such attacks involves complex privacy-utility considerations

[Fredrikson15]

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

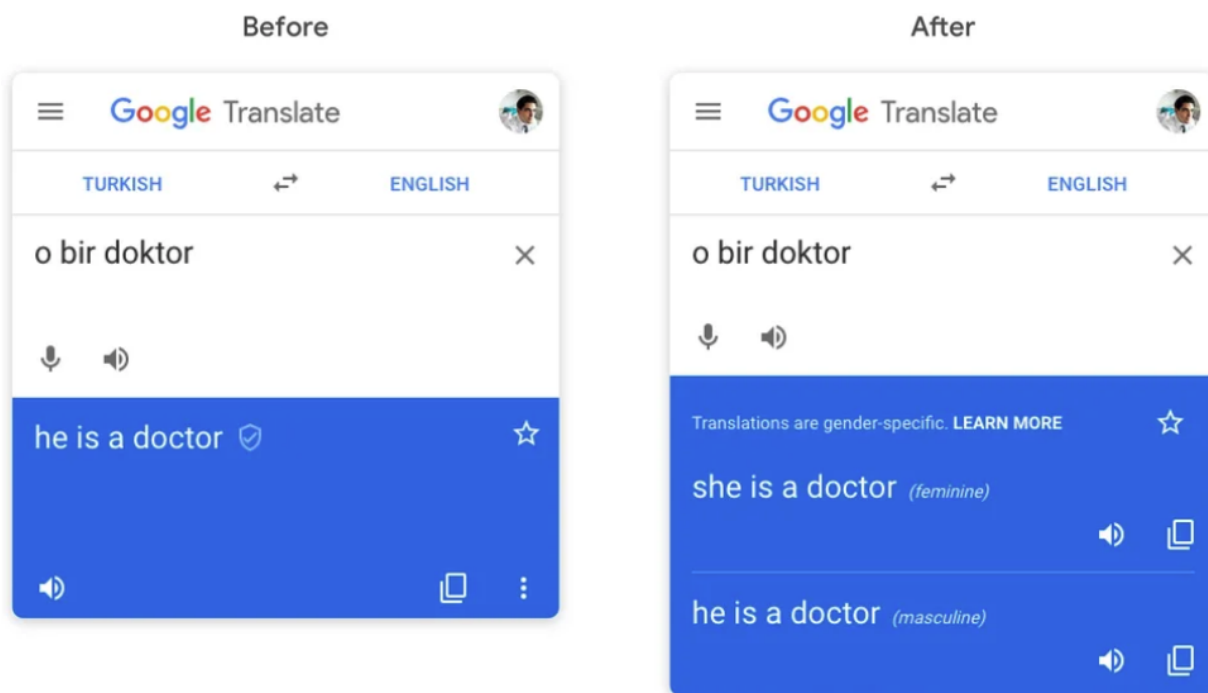
Why Ethic and Fairness in Machine Machine Learning?



--> Law Alone is not Sufficient

Example: Gender Bias in Google Translator

- Sentences from gender-neutral languages like Farsi or Turkish resulted in gender stereotypical translations



Example: Gender Bias in Google Translator

- Reasons for gender bias
 - Ordering (social order bias)
 - Biased descriptions
 - Metaphors
 - Presence of women in text
 - Men often described based on behavior
 - Women are described based on their sexuality and external appearance

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

1. Regulatory Challenges in Machine Learning
2. Privacy-Preserving Machine Learning Techniques: Applications and Limitations
3. Fairness in Machine Learning Models for Credit Scoring and Loan Approval
4. Data Bias in AI Healthcare Applications: Algorithmic Fairness in Diagnosis and Treatment
5. Algorithmic Bias in Hiring and Recruitment: Examining Fairness and Ethics
6. An Analysis of Individual Fairness and Group Fairness

[T1] Regulatory Challenges in Machine Learning

- Expected results: A detailed analysis of regulation, their impact on the market and gaps in the legal framework.
 - How effective are existing regulations, such as GDPR and the EU AI Act, in governing machine learning systems?
 - What are the gaps in current legal frameworks, and what additional regulations might be needed to address challenges?
 - How do different global regulations impact cross-border applications of machine learning models?
 - ...

- Expected results: A comprehensive evaluation of privacy-preserving machine learning techniques and their effectiveness in safeguarding user data.
 - How do privacy-preserving techniques (Differential Privacy, Homomorphic Encryption, Federated Learning, Secure Multiparty Computation) affect the balance between keeping data safe and having accurate machine learning models?
 - How well do current methods measure the privacy protection provided by different techniques?
 - ...

- Expected results: A detailed analysis of issues in credit scoring and loan approval using ML.
 - What are the problems that unfair decisions and are there unique factors in the financial industry?
 - Are the current regulations sufficient to prevent bias in credit scoring and loan approval?
 - How accurate are credit scoring and loan approval models using ML compared to traditional techniques?
 - ...

[T4] Data Bias in AI Healthcare Applications: Algorithmic Fairness in Diagnosis and Treatment

- Expected results: A detailed analysis of ethic issues that result from using AI in healthcare applications.
 - Where is AI used in health care?
 - How to balance the benefits of AI healthcare applications with the need to ensure ethical and responsible use?
 - Do legal frameworks and guidelines exist already and what are their recommendations?
 - ...

- Expected results: A detailed analysis of Algorithmic Bias in Hiring and Recruitment.
 - What are the causes of bias in AI hiring and how does it affect job seekers?
 - How can AI hiring/recruitment systems be designed to ensure fairness and ethics?
 - Do regulatory frameworks already exist to address AI recruitment bias, and how effective are they?
 - ...

[T6] An Analysis of Individual Fairness and Group Fairness

- Expected results: A detailed analysis of individual fairness concepts.
 - What definitions of individual fairness and group fairness exist?
 - Are there issues in the different fairness definitions?
 - How does individual fairness differ from group fairness
 - ...

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



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

- [Cavoukian2010]: Privacy by Design The 7 Foundational Principles Implementation and Mapping of Fair Information Practices, 2010.
- [D' Acquisti2015]: 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_warf2.htm
- [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).
- [Dwivedi 23] Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71, 102642.
- [Leavy 18] Leavy, S.: Gender bias in artificial intelligence: the need for diversity and gender theory in machine learning. In: Proceedings of the 1st International Workshop on Gender Equality in Software Engineering, pp. 14–16 (2018)



Chair of Mobile Business & Multilateral Security

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

