

# Data Privacy Impact Assessment (DPIA) and Fundamental Rights Impact Assessment (FRIA)

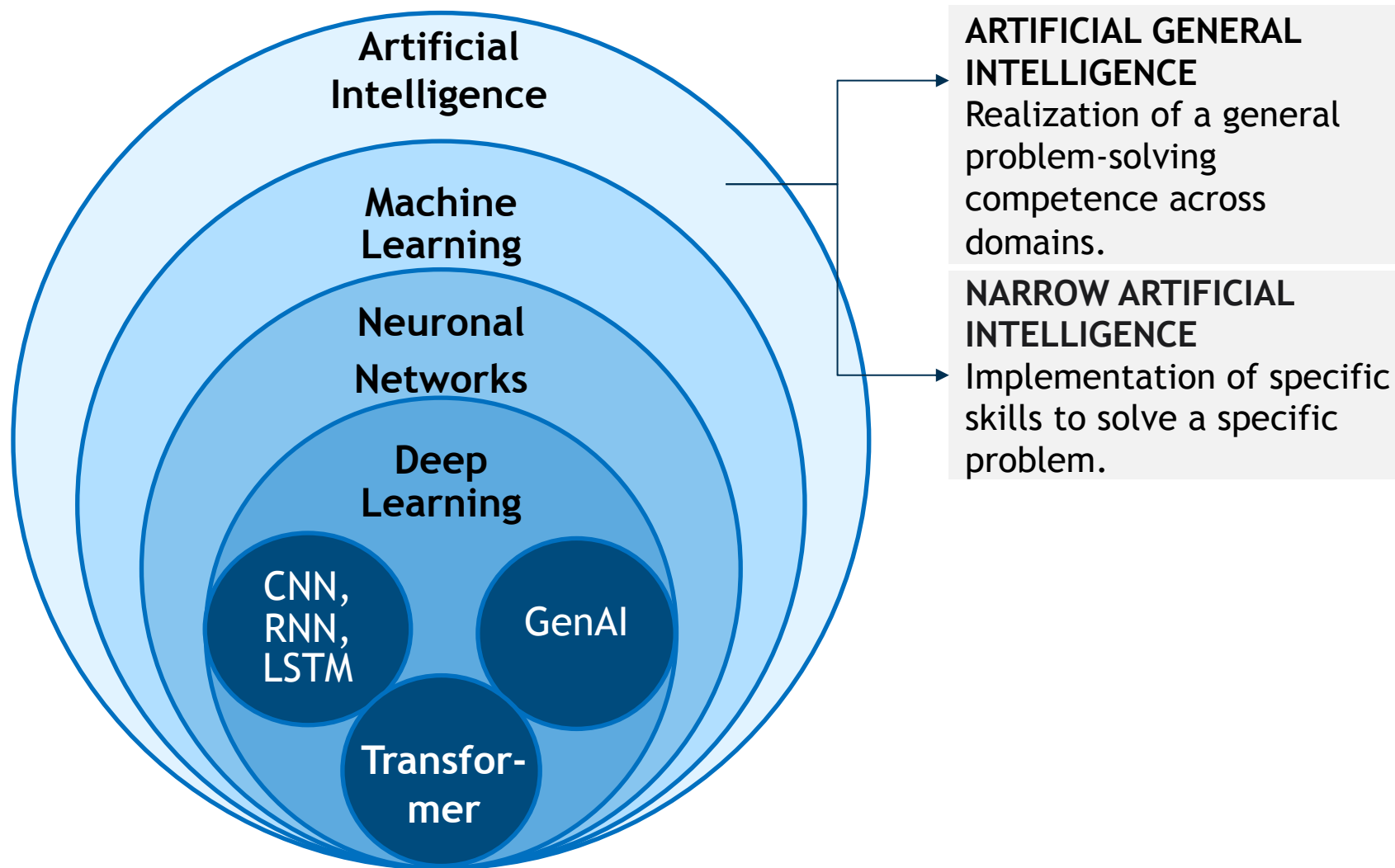
**Sascha Löbner**

loebner-science.de

Chair of Mobile Business & Multilateral Security

Goethe University Frankfurt

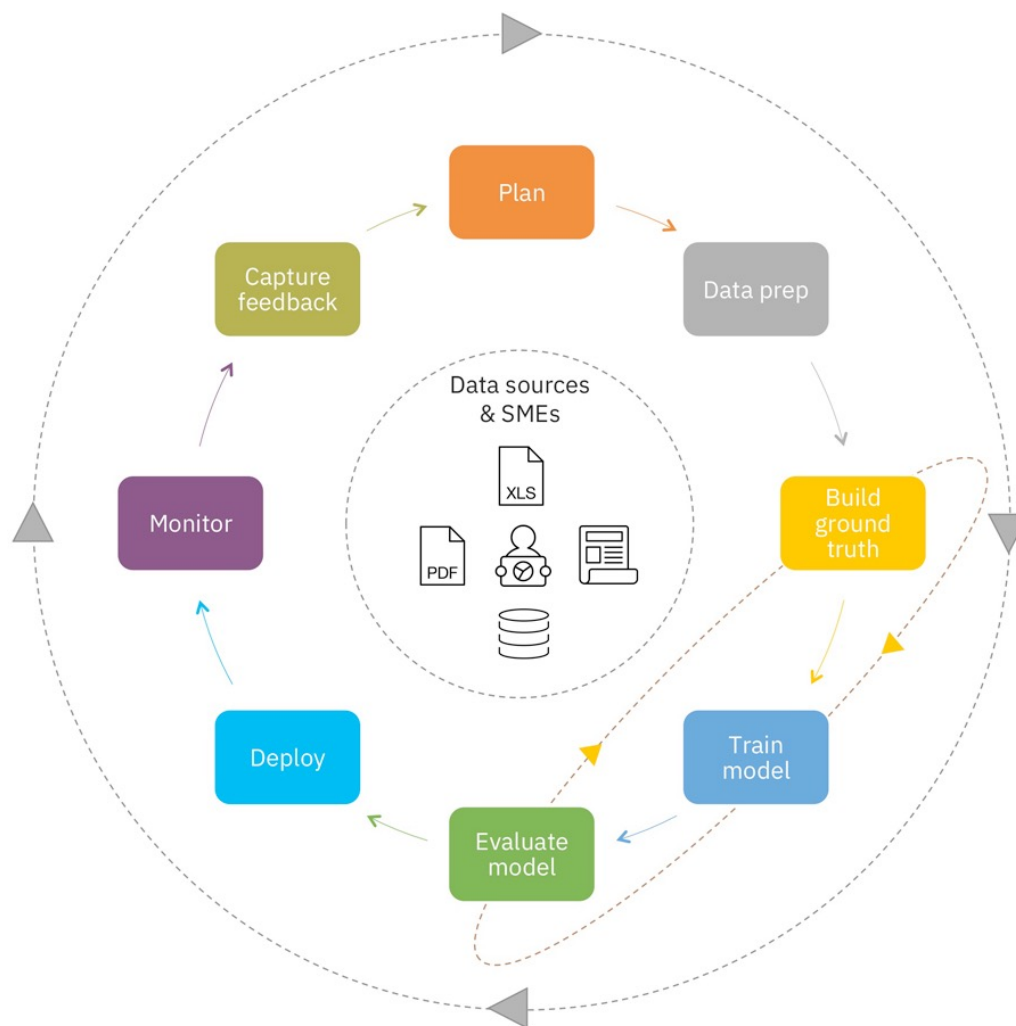




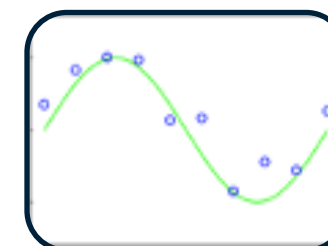
- **“Machine learning** is defined as an automated process that extracts patterns from data” [Kelleher2015]
  - **Supervised Learning:** Applications in which the training data comprises examples of the input vectors along with their corresponding target vectors [Bishop2006].
  - **Unsupervised Learning:**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

# Machine Learning Life Cycle

## CRISP-DM - additional steps by IBM

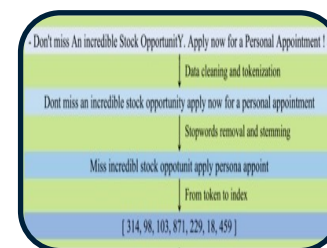


- **Clustering:** Dividing the dataset into clusters of similar examples. (e.g. spam filter)
- **Classification:** The computer program is asked to specify which of  $k$  categories some input belongs to. (e.g. object recognition)
- **Regression:** The computer program is asked to predict a numerical value given some input. (e.g. price prediction)

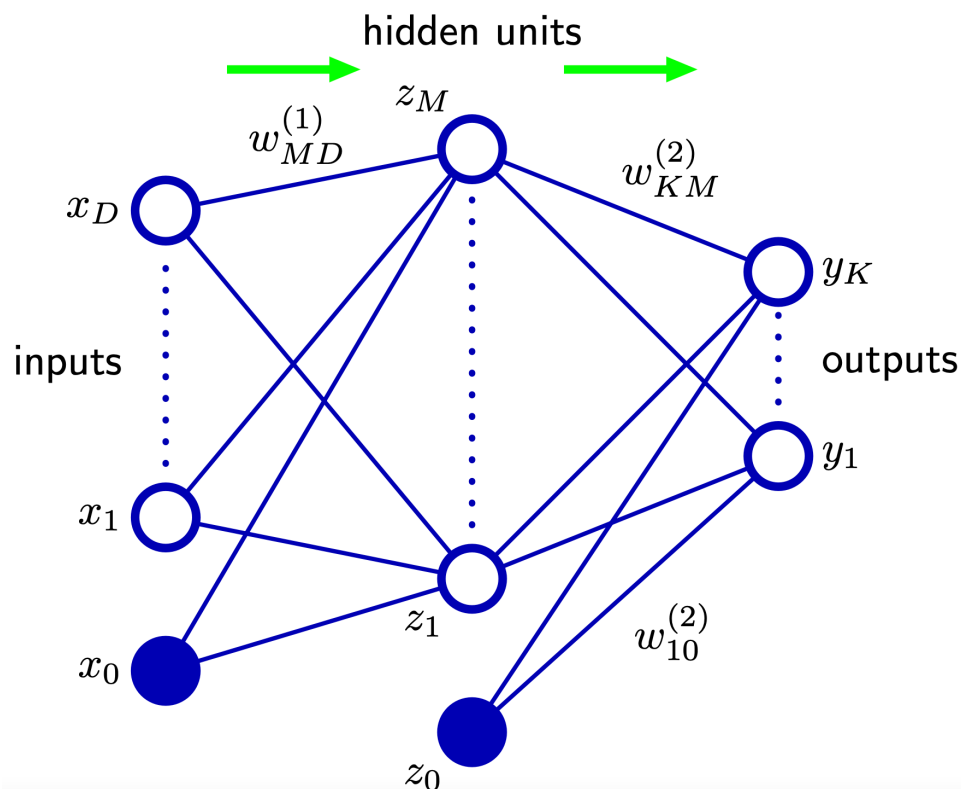


[Goodfellow2016]

- **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]

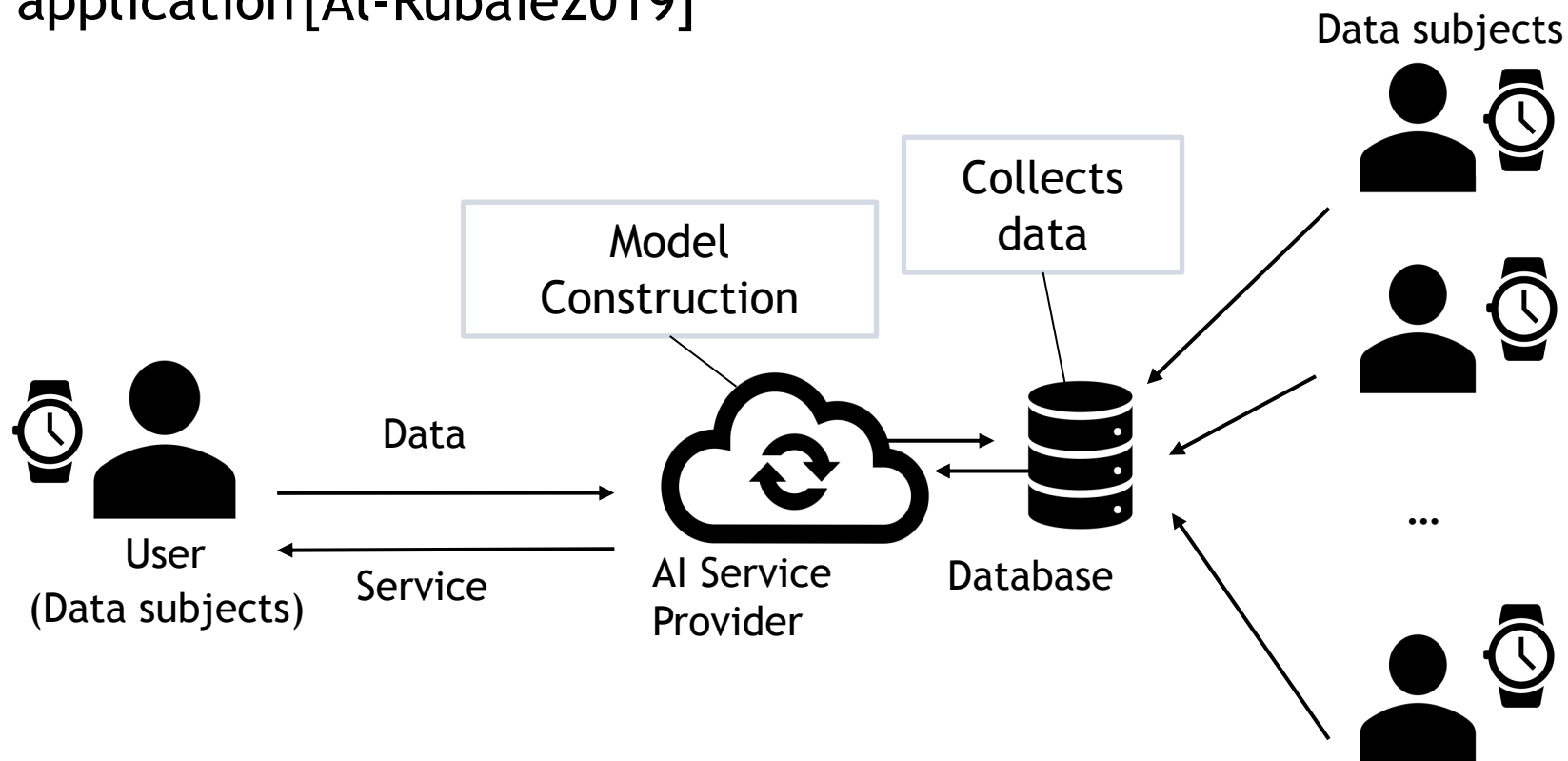
**“ChatGPT:** Uses a combination of unsupervised pre-training and supervised fine-tuning to generate human-like responses to queries and provide responses to topics that resemble that of a human expert.”



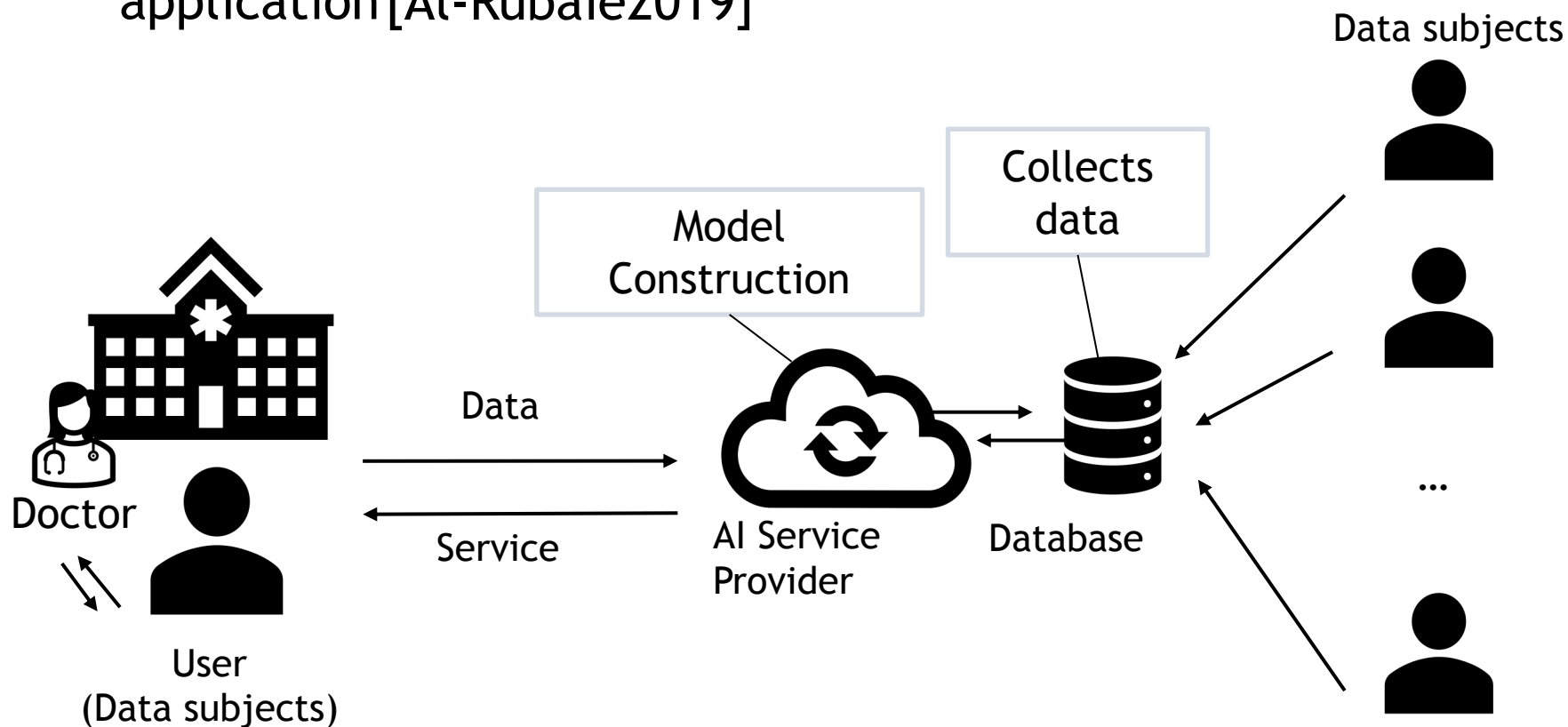
# Risks in Machine Learning

# Example Fitness Watch

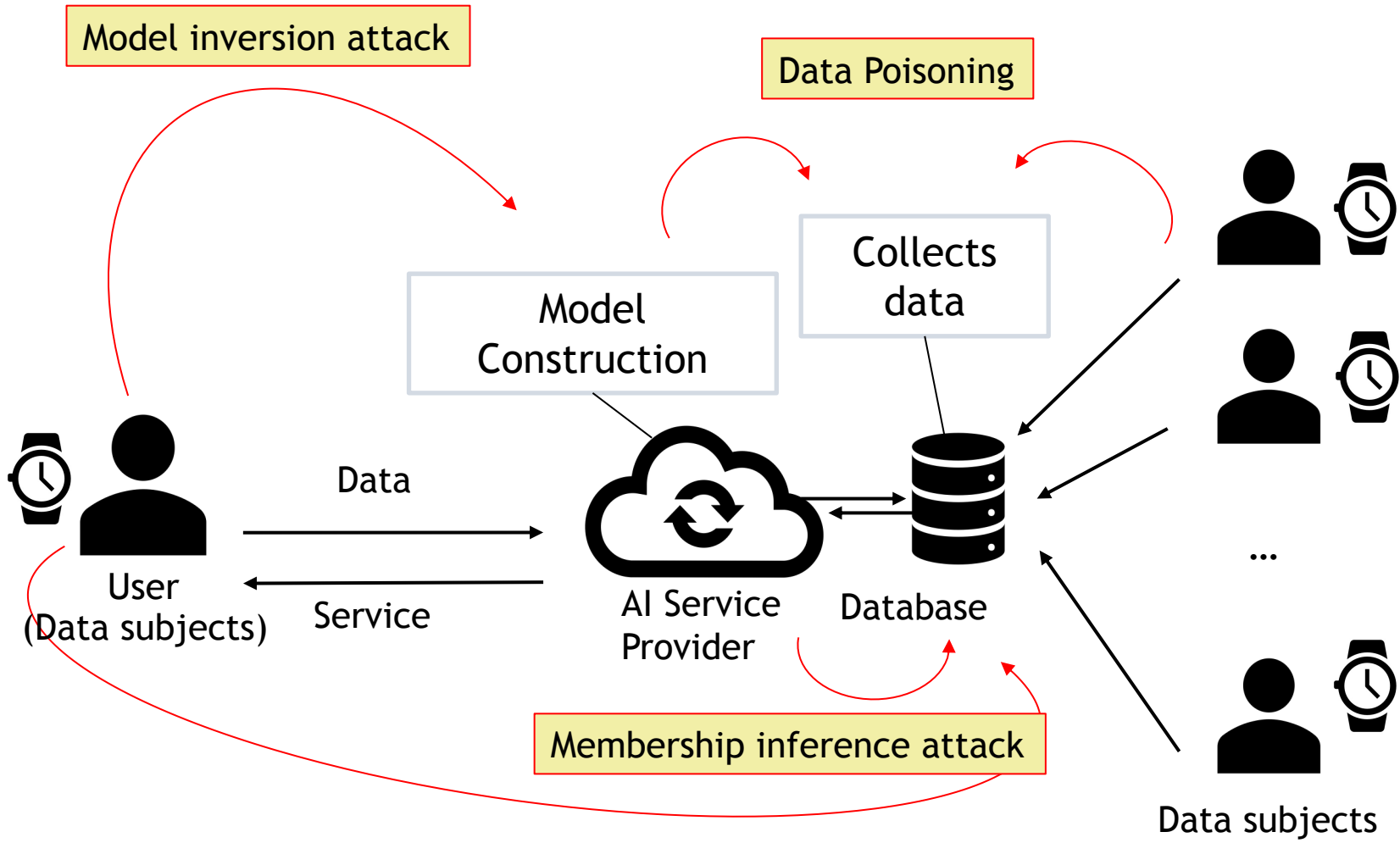
- Machine Learning (ML) is becoming part of our daily life
- Often individuals' private data is required for the ML application [Al-Rubaie2019]



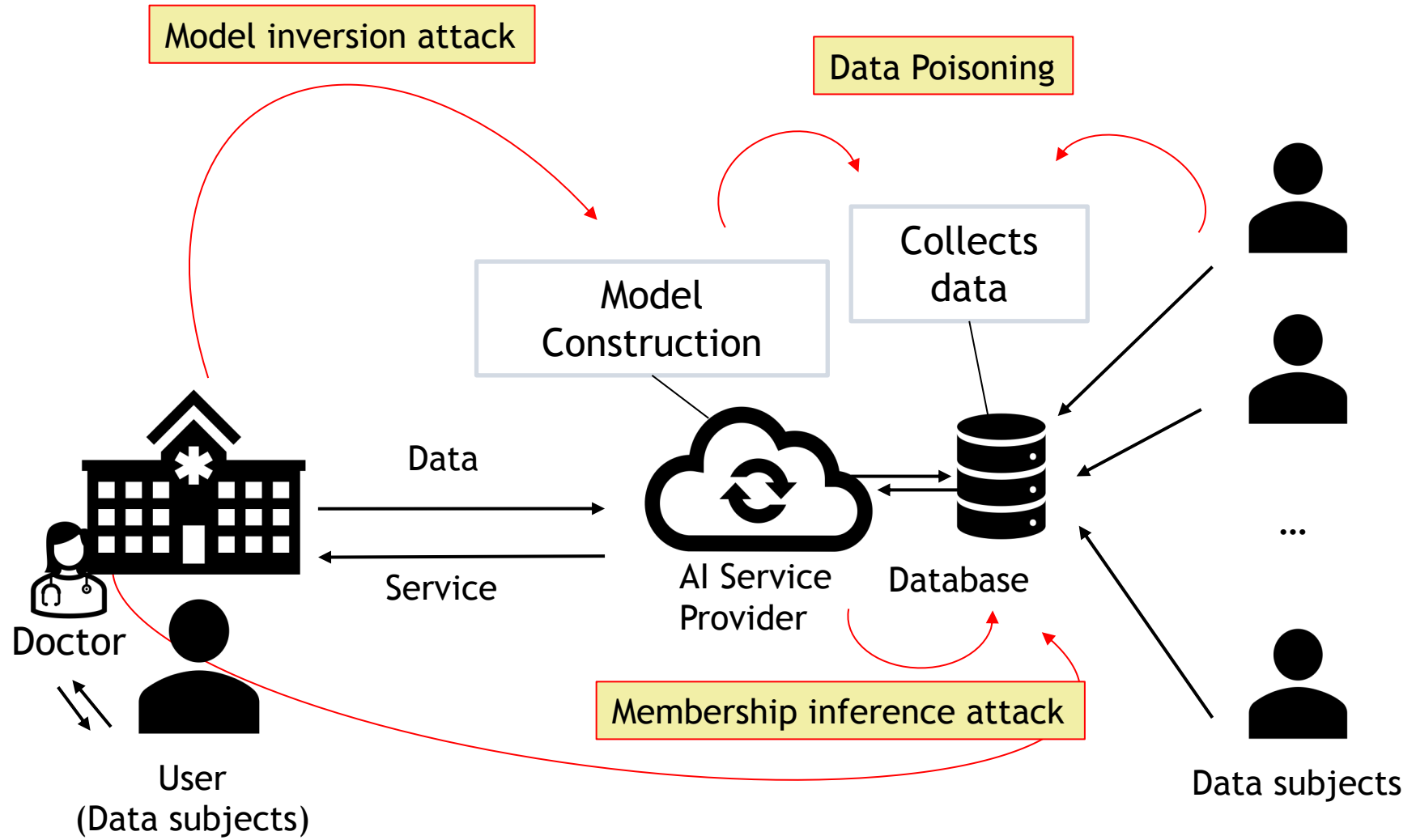
- Machine Learning (ML) is becoming part of our daily life
- Often individuals' private data is required for the ML application [Al-Rubaie2019]



# 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

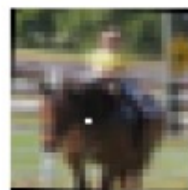
Example: One pixel attack:

- 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

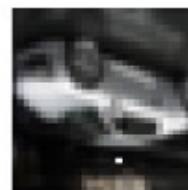
**AllConv**



**SHIP**  
**CAR(99.7%)**



**HORSE**  
**DOG(70.7%)**



**CAR**  
**AIRPLANE(82.4%)**

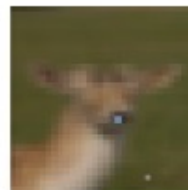
**NiN**



**HORSE**  
**FROG(99.9%)**



**DOG**  
**CAT(75.5%)**

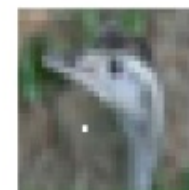


**DEER**  
**DOG(86.4%)**

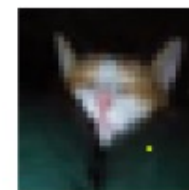
**VGG**



**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]



## DPIA and FRIA

# DPIA: Legal Basis and Purpose

- The legal mandate for the DPIA is set out in Article 35 of Regulation (EU) 2016/679 (GDPR): a DPIA is defined as a process designed to achieve specific goals related to managing the risks inherent in processing personal data.
- The process requires the controller to:
  - Describe the processing operations envisaged, systematically.
  - Assess the necessity and proportionality of the processing in relation to its purposes.
  - Help manage the risks to the rights and freedoms of natural persons resulting from the processing of personal data. This involves assessing risks and determining measures to address them.
- The DPIA is an important tool for accountability as it assists controllers not only in complying with the GDPR requirements but also in demonstrating that appropriate measures have been taken to ensure compliance with the Regulation.
- A DPIA is **not mandatory** for **every** processing operation.
- The controller is the party obliged to carry out the DPIA and must seek the advice of the Data Protection Officer (DPO), where one is designated.
- Failure to adhere to DPIA requirements can result in significant administrative fines.

# DPIA Triggers: Indicators of High Risk

A DPIA is only required where a type of processing is “likely to result in a high risk to the rights and freedoms of natural persons” (Article 35(1) GDPR). Examples:

- Processing personal aspects based on automated means, including profiling, leading to decisions with legal or similarly significant effects.
- Processing on a large scale (factors include the number of data subjects, volume of data, duration, geographical extent) of special categories of data (Art. 9 GDPR) or data relating to criminal convictions (Art. 10 GDPR).
- Systematic monitoring of a publicly accessible area on a large scale.
- Use of a new technological solution (e.g., AI systems), as this can involve novel forms of data collection and usage.
- Processing data concerning vulnerable individuals (e.g., children, patients).
- Operations that combine datasets from different purposes/controllers, exceeding reasonable expectations.

## However:

- A DPIA is **not mandatory** for every processing operation.
- However, even when not legally required, a DPIA is often **recommended** as a useful tool for compliance.

# DPIA Scope and Timing

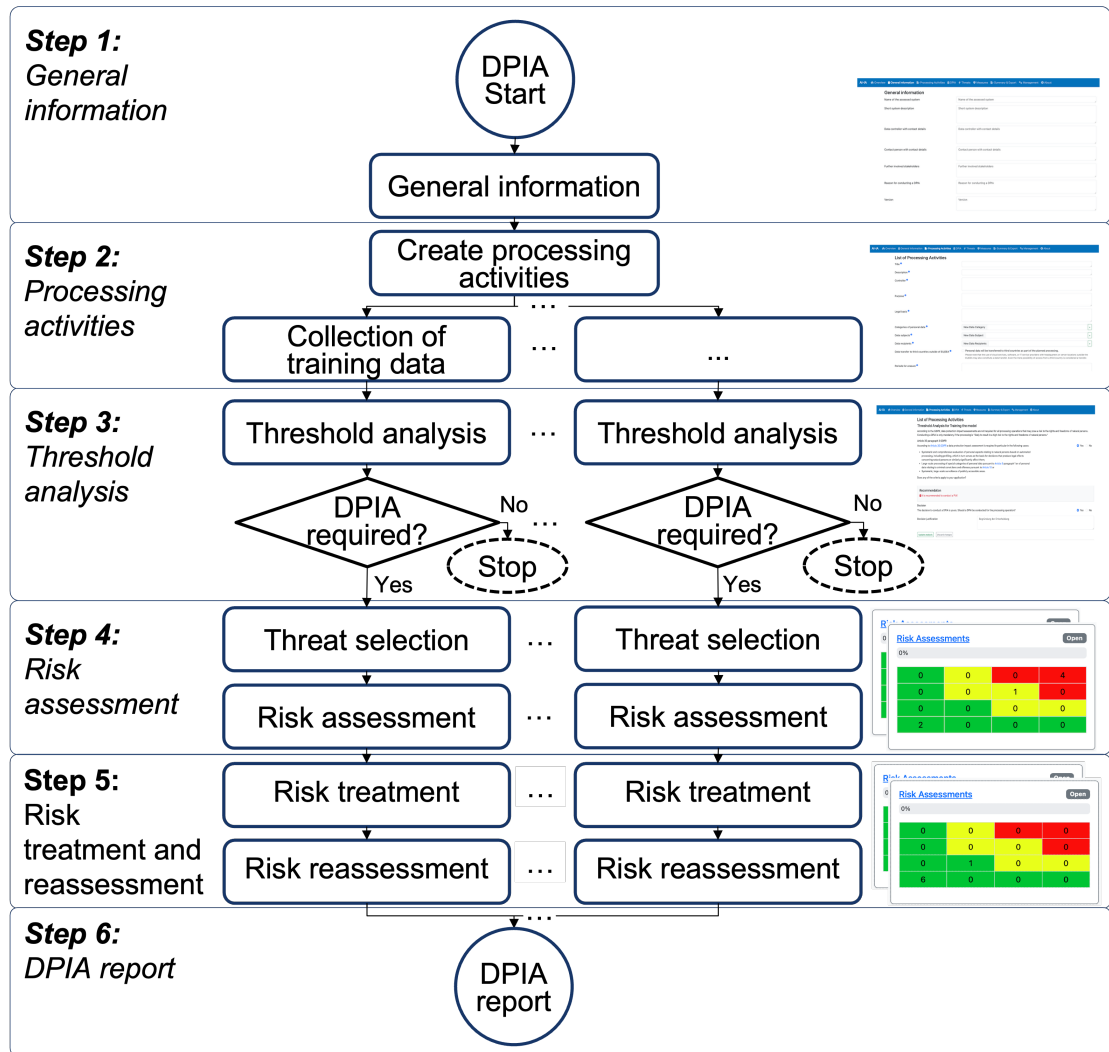
- A DPIA is often tailored to a **single processing operation**. However, for efficiency and practicality, the GDPR permits a broader scope under specific conditions:
  - A single assessment may address a set of similar processing operations that present similar high risks. This allowance is reflected in GDPR Article 35(1).
  - The similarity criteria mean that multiple processing operations can be covered by a single DPIA if they are similar in terms of their nature, scope, context, purpose, and the resulting risks.
  - **Example:** a single hospital operator covering video surveillance in all its hospitals, could carry out a single DPIA.
- The timing of the DPIA is critical to fulfilling the principle of “data protection by design” (Art. 25 GDPR).
- The DPIA must be carried out “prior to the processing” (Art. 35(1) and (10) GDPR (and Recitals 90, 93)).
- **Ideally**, the DPIA should be started as early as is practicable in the design of the processing operation, **even if some details are still unknown**. Updating the DPIA throughout the project lifecycle ensures that data protection and privacy are continuously considered.
- **However:** the fact that the DPIA might need subsequent updates is not a valid reason for postponing or not carrying out a DPIA initially.

# DPIA Minimum Content (Article 35(7) GDPR)

- The GDPR, specifically in Article 35(7), specifies the minimum elements that a DPIA **must** contain. Meeting these requirements is crucial for the controller to comply with the regulation and demonstrate accountability.
- The assessment must contain at least:
  1. A systematic description of the processing and purposes (Article 35(7)(a)).
  2. An assessment of the necessity and proportionality of the processing (Article 35(7)(b)).
  3. An assessment of the risks to the rights and freedoms of data subjects (Article 35(7)(c)).
  4. Measures envisaged to address the risks, including safeguards and security measures (Article 35(7)(d)).
- The DPIA functions as a systematic risk management tool that prioritizes the impact on individuals.
  - The risk management inherent in a DPIA aims at “managing risks” to the rights of the data subjects, and thus takes their perspective. This **contrasts with risk management in other fields, such as information security**, which often focuses solely on the organization's needs.
  - The reference to the “rights and freedoms” of data subjects primarily concerns the rights to data protection and privacy, but may also involve other fundamental rights such as freedom of speech or non-discrimination.

# Prior Consultation for High Residual Risks

- The relationship between **risk assessment** and **regulatory involvement** is critical, especially when risks cannot be fully mitigated.
  - The process involves determining if the identified risks can be considered acceptable in view of the existing or planned controls. If they are acceptable, the processing can proceed without consulting the supervisory authority.
  - However, if the DPIA reveals high residual risks, meaning the risks that remain after the controls and technical and organizational measures have been implemented and these risks **cannot be sufficiently addressed or reduced** by the data controller, consultation with the supervisory authority is mandatory.
- This requirement is known as **Prior Consultation**, mandated by Article 36(1) GDPR.
- Consultation is necessary in cases where data subjects may encounter **significant, or even irreversible**, consequences which they may not overcome (e.g., identity theft following disclosure of data, or a threat to life) and/or when it seems obvious that the risk will occur.
- **If** the DPIA reveals high residual risks, the controller must fully provide the DPIA report to the supervisory authority. Failure to consult the authority when required can result in administrative fines. The supervisory authority *may* then issue an opinion stating that the planned processing is in breach of the GDPR if it proves impossible to sufficiently reduce the risks.



# FRIA: Context and Legal Basis

- The FRIA is introduced by the EU Artificial Intelligence Act (AI Act), officially Regulation (EU) 2024/1689. The specific obligations regarding the FRIA are detailed in **Article 27** of this Regulation.
- The overarching objective of the AI Act is to improve the functioning of the internal market by establishing a uniform legal framework that promotes the uptake of human-centric and trustworthy AI.
- The specific aim of the FRIA is for the deployer to proactively identify the specific risks to the rights of individuals or groups of individuals that the high-risk AI system may present. The FRIA also helps to “identify measures [to take] in the case of a materialization of those risks”.
- A FRIA must be carried out for specific high-risk AI systems. The AI Act adopts a **risk-based approach**, classifying systems as high-risk if they can have a significant harmful impact on the health, safety, and fundamental rights of persons in the EU.
- Organizations must conduct the FRIA “prior to deploying the high-risk AI system”. It is not a one-time step; deployers are required to update the FRIA when they consider that any of the relevant factors have changed.
- The FRIA obligation applies to specific deployers of high-risk AI systems referred to in Article 6(2) of the AI Act (Annex III systems), with the exception of those listed in point 2 of Annex III (Critical infrastructure).

# When is a FRIA Mandatory?

- A FRIA is required prior to deploying a high-risk AI system.
- Systems are classified as high-risk if they are considered to have a significant harmful impact on the health, safety and fundamental rights of persons in the EU. The extent of the adverse impact on fundamental rights (e.g., human dignity, non-discrimination, data protection, right to an effective remedy) is particularly relevant in this classification .
- AI systems are categorized as high-risk under two main criteria:
  - 1. Safety Components of Products:** The AI system is a safety component of a product, or the product itself, covered by specific Union harmonization legislation (Annex I of the AI Act), and that product requires third-party conformity assessment.
  - 2. Specific Use Cases (Annex III):** The system's intended purpose falls within the categories listed in Annex III of the AI Act.

*Note on Derogation (Article 6(3)): Even if a system falls under Annex III, it is not considered high-risk if it does not pose a significant risk of harm to health, safety, or fundamental rights, such as if it only performs a narrow procedural task or a task preparatory to an assessment. However, an AI system is always considered high-risk if it performs profiling of natural persons.*

# Who Conducts the FRIA?

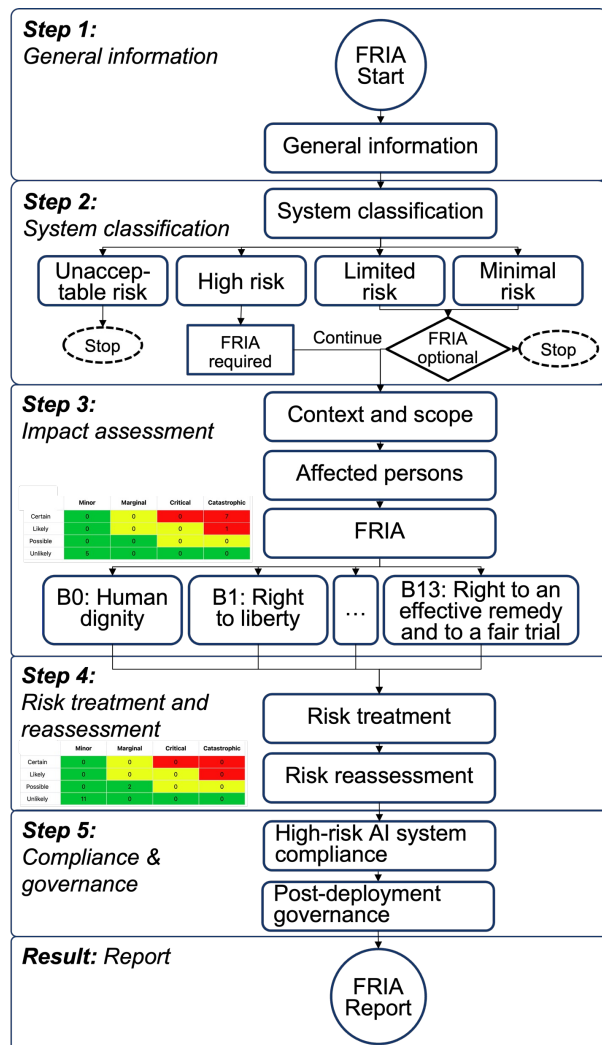
- The FRIA must be performed by the Deployer of the high-risk AI system.
- This typically includes:
  1. Bodies governed by public law (public authorities/agencies).
  2. Private entities providing public services.
  3. Deployers of specific high-risk AI systems listed in Annex III of the AI Act.
- The aim is for the deployer to proactively identify specific risks to the rights of individuals or groups and determine remediation strategies.
- A “deployer” is the natural or legal person using an AI system under its authority.
  - The AI Act defines a “deployer” in Article 3, point (4) as: “a natural or legal person, public authority, agency or other body using an AI system under its authority”.
  - Exception: the AI system is used “in the course of a personal non-professional activity”.
  - Essentially, a deployer is the end-user entity, whether public or private that puts an AI system into operation for a specific application or purpose within its activities

# FRIA Core Content (Article 27(1) AI Act)

The FRIA must detail the impact on fundamental rights and include:

- A description of the deployer's processes where the AI system will be used.
- The categories of natural persons and groups likely to be affected.
- The specific risks of harm likely to impact fundamental rights.
- The implementation of human oversight measures.
- Measures to be taken in case the risks materialize (including internal governance and complaint mechanisms).
- The FRIA must take into account information provided by the AI system provider.
- Deployers must notify the market surveillance authority of the FRIA results.

# Our Approach to FRIA



# The Complementary Role of DPIA and FRIA

- The relationship between the DPIA (under GDPR) and the FRIA (under the AI Act) is fundamentally complementary. This relationship is **explicitly anchored in the AI Act** to ensure that existing accountability efforts are leveraged while new, broader obligations are met.
- Where DPIA obligations (Art. 35 GDPR) **are already met**, the FRIA (Art. 27 AI Act) shall **complement** that DPIA.
- Organizations deploying high-risk AI systems that process personal data will need to conduct **both**.
- Both the DPIA and the FRIA share the common methodological feature of being risk-based assessment tools that must be performed prior to engaging in high-risk activities.
  - DPIA (GDPR): The assessment is mandatory where a type of processing is “likely to result in a high risk to the rights and freedoms of natural persons” (GDPR Article 35(1)). The focus primarily concerns the rights to data protection and privacy, but may also involve other fundamental rights such as freedom of speech, prohibition of discrimination, or right to liberty.
  - FRIA (AI Act): The objective is to evaluate the risks to the rights of individuals or groups of individuals that the high-risk AI system may present.
- The AI Act’s FRIA extends beyond data protection to cover broader fundamental rights, health, and safety risks.

- [Tolpegin20] Vale Tolpegin, Stacey Truex, Mehmet Emre Gursoy, and Ling Liu. Data poisoning attacks against federated learning systems. In European Symposium on Research in Computer Security, pages 480–501. Springer, 2020.
- [Su19] Su, J., Vargas, D. V., & Sakurai, K. (2019). One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation*, 23(5), 828-841.
- [weforum19]: World Economic Forum, <https://www.weforum.org/publications/the-next-generation-of-data-sharing-in-financial-services-using-privacy-enhancing-techniques-to-unlock-new-value/> [accessed 14.01.2024]
- [Lin16] Lin, C., Song, Z., Song, H., Zhou, Y., Wang, Y., & Wu, G. (2016). Differential privacy preserving in big data analytics for connected health. *Journal of medical systems*, 40, 1-9.
- [Al-Rubaie18] Al-Rubaie, M., & Chang, J. M. (2019). Privacy-preserving machine learning: Threats and solutions. *IEEE Security & Privacy*, 17(2), 49-58.
- [Kaissis20] Kaissis, G. A., Makowski, M. R., Rückert, D., & Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, 2(6), 305-311.
- [Yang19] Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2), 1-19.
- [Fredrikson15] Fredrikson, M., Jha, S., & Ristenpart, T. (2015, October). Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security* (pp. 1322-1333).
- [IBM17], Adapt DevOps to cognitive and artificial intelligence systems, <https://web.archive.org/web/20190503084621/https://developer.ibm.com/articles/cc-devops-artificial-intelligence-cognitive/> [accessed 03.05.2019]
- [Wirth00] Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 1, pp. 29-39).



## Chair of Mobile Business & Multilateral Security

**Prof. Dr. Kai Rannenberg**  
Goethe University Frankfurt  
E-Mail: [kai.rannenberg@m-chair.de](mailto:kai.rannenberg@m-chair.de)  
WWW: [www.m-chair.de](http://www.m-chair.de)

