The Answers to The Questions

1. Name five different attribution methods.

In particular, there are the list of five perturbation-based methods:

1. SARFA (explains the actions of RL agents in board games and Atari games (RL entities)

2. NLIZE(NLP - qualitative analysis)

3. MFPP (images - pointing games)

4. Temporal Masks (video)

5. Auto Focus (software code - qualitative analysis)

2. Name five different methods applicable to privacy-preserving deep learning.

In particular, there are the list of five methods:

1. MiniONN (NN, PaaS)

2. DeepSecure (CNN, Cloud Service)

3. E2DM

4. PATE (GAN)

5. CryptoDL

3. Explain in your own words why privacy-preserving and XAI methods follow competing goals.

The use of artificial intelligence in medicine and healthcare has led to successful clinical applications in several areas. The conflict between the requirements for the use of data and the protection of confidentiality in such systems must be resolved in order to achieve optimal results, as well as compliance with ethical and legal norms. This requires innovative solutions such as privacy-preserving machine learning (PPM). The consequence of this is a high probability that the calculations based on data without disclosing the original content will be impossible sometimes. That is why specialists make calculations based on data without revealing the original content. Such models are quite accurate in their predictions and have rathe high security level, but it is difficult for experts using such systems to understand how this system makes a decision. For the same reason, there are difficulties with configuring such systems for specific parameters, so the system becomes quite rigid and does not take into account the specifics of a particular client's activity and critical to it’s security. There is a so-called "black box" problem.

The solution to the above problem is the use of methods Explainable Artificial Intelligence (XAI), which results of the solution can be understood by people. This is based on the "white box" concept and contrasts with the "black box" problem of PPM. Thus, privacy-preserving and XAI methods follow competing goals.