Lernende Systeme
GERMANY'S PLATFORM FOR ARTIFICIAL INTELLIGENCE

Distributed Machine Learning

Improved data protection for AI applications?

AT A GLANCE Distributed Machine Learning

- promises improved data protection by design and higher performance.
- trains Machine Learning (ML) models decentrally on end devices instead of centrally on a server.
- uses edge computing for AI and distributes the computing load.
- enables – often personal – training data to remain on end devices and thus with the the users.
- can use this data sovereignty to ensure the protection of personal data and increase informational self-determination.
- is already in use in the federated learning variant; other approaches are still at the research stage or on the threshold of market entry.
- can be used in a variety of ways, such as for mobility or health applications.

However, distributed Machine Learning creates new gateways for attackers and potentially creates a deceptive sense of security. Some experts therefore warn against exaggerated expectations in terms of data protection.

Starting Point

AI systems are based on training with large amounts of – sometimes sensitive – data. The use of this data is sometimes in tension with data protection and the individual's right to decide for himself or herself on the disclosure and use of personal data (informational self-determination). This case, for example, when an AI system only makes certain suggestions to users based on their search history and hides others that may be more suitable. At the same time, there are legal uncertainties for companies when training AI systems: According to the General Data Protection Regulation (GDPR), personal data may generally only be used for a specific purpose; for other purposes it may be necessary to obtain the subsequent consent of these persons or to balance the individual interests. The latter is complex and open to interpretation.

However, there are technical approaches that effectively combine data use and data protection – and may create new market opportunities for privacy-preserving AI solutions. These include the approach of distributed Machine Learning.