KI und Datenschutz: Large Language Models und Data Sharing

Die Internationale Arbeitsgruppe für Datenschutz in der Technologie (IWGDPT) – auch genannt „Berlin Group" – hat am 27. Dezember 2024 unter dem Vorsitz der Bundesbeauftragten für Datenschutz und Informationsfreiheit zwei wegweisende Arbeitspapiere zum Thema "Large Language Models" (LLMs) und "Data Sharing" veröffentlicht.

Um Innovationen voranzutreiben, neue Erkenntnisse zu gewinnen und neue Produkte und Dienstleistungen zu entwickeln, kommt es immer häufiger zu „Data Sharing" zwischen Organisationen. Darunter wird die Bereitstellung von Daten zur Nutzung durch andere verstanden. Es findet also eine Datenübertragung zwischen Organisationen statt. Unter Umständen kann das Teilen von Daten bzw. der Austausch von Daten, beispielsweise aufgrund mangelnder Transparenz oder fehlender Sicherheitsvorkehrungen, zu größeren Datenschutzverstößen und Datenmissbrauch führen. Die Berlin-Group hat sich dieser Problematik in ihrem Papier zum „Data-Sharing" angenommen und Empfehlungen bzw. Leitlinien für die verschiedenen Akteure des Datenaustausch entwickelt, um eine vertrauenswürdige Umgebung für Datenaustausch zu schaffen. Dabei sollen Datenschutzgrundsätze gewahrt bleiben und gleichzeitig das Potenzial eines sicheren und geschützten Datenaustauschs maximiert werden.

Zur Generierung von Texten werden oftmals sog. Large Language Models (kurz LLMs) eingesetzt. Damit das gelingt werden diese vorab mit großen Datenmengen trainiert – sind darunter personenbezogene Daten, kann es unter Umständen auch hier zu Datenschutzverstößen kommen. In ihrem zweiten Papier zu „Large Language Models" gibt die Berlin Group einen umfassenden Überblick darüber, wie ein datenschutzkonformer Einsatz derer gelingen kann.

Data-Sharing

Weltweit gilt es bereits zahlreiche gesetzgeberische Initiativen, die das Data Sharing im öffentlichen als auch im privaten Sektor regulieren. Beispiele hierfür sind der Data Governance Act und der Data Act auf EU-Ebene, die Regeln für den Austausch von Daten festlegen und den Zugang zu Daten ermöglichen. Dennoch birgt der Austausch von Daten immer das Risiko von Datenmissbrauch und Datenschutzverstößen.

Um das zu vermeiden, wird in dem Papier zum einen empfohlen, Privacy-Enhancing Technologies (PETs) zu implementieren. Um Datenschutzvorgaben einzuhalten, sollte zudem eine exakte Analyse darüber stattfinden, warum und an wen Daten geschickt werden. Zudem enthält das Papier Empfehlungen für verschiedene Akteure wie Datenverantwortliche, Gesetzgeber, Technologieanbieter, Forschungsgemeinschaft und Datenschutzbehörden. Betroffene sollten ausreichend über den Datenaustausch informiert werden und müssen die Möglichkeit haben einer solchen Datenübermittlung widersprechen zu können.

Large-Language-Models

Künstliche Intelligenz, die Texte in menschlicher Sprache generieren kann, basieren meist auf einem LLM. Die Grundlage für diese KI-Modelle bilden meist große Datenmengen, mit denen die LLMs vorab trainiert werden. In dem Papier der Berlin Group zu diesen KI-Modellen werden die Risiken solcher KI-Modelle für den Datenschutz aufgezeigt. So kann es durch die Nutzung von LLMs beispielsweise dazu kommen, dass Desinformationen verbreitet werden, da die LLMs fehlerhafte Inhalte generieren oder Betroffene die Kontrolle über ihre Daten verlieren, da unklar ist, inwiefern die Daten genutzt werden.

Um das zu vermeiden, betont die Arbeitsgruppe immer wieder die Bedeutung hochwertiger und datenschutzkonformer Trainingsdaten. Zudem soll neben den allgemeinen Prinzipien des Datenschutzes wie Transparenz, Zweckbegrenzung und Erlaubnisgrundlage für die Datenverarbeitung, auch daran gearbeitet werden, die Verarbeitung von personenbezogenen Daten durch LLMs möglichst zu vermeiden.

Gleichzeitig werden technische Designs der KI-Modelle vorgeschlagen, um Datenschutzgrundsätze technisch zu ermöglichen. Durch bspw. Differential Privacy, kann die Wahrscheinlichkeit reduziert werden, dass personenbezogene Daten rekonstruiert werden oder durch Machine Unlearning, welches es erlaubt, Daten nachträglich aus einem Modell zu entfernen. Dadurch wird deutlich, dass der erfolgreiche und datenschutzkonforme Einsatz von LLMs auch innovative technische Maßnahmen gewährleistet werden.

Für alle weiteren Fragen rund um das Datenschutzrecht stehen Ihnen gerne zur Verfügung

Dr. Kristina Schreiber
+49 221 65065-337
kristina.schreiber@loschelder.de

Dr. Simon Kohm
+49 221 65065-200
simon.kohm@loschelder.de

Dennis Pethke, LL.M.
+49 221 65065-337
dennis.pethke@loschelder.de

Rebecca Moßner
+49 221 65065-465
rebecca.mossner@loschelder.de

Impressum

LOSCHELDER RECHTSANWÄLTE
Partnerschaftsgesellschaft mbB
Konrad-Adenauer-Ufer 11
50668 Köln

Tel. +49 (0)221 65065-0, Fax +49 (0)221 65065-110
info@loschelder.de
www.loschelder.de