

Rechtlicher Schutz der beim maschinellen Lernen verwendeten KI Trainingsdaten – Flucht in den Know-How Schutz bei KI Entwicklungen?

Prof. Dr. Nils Heide¹

Das maschinelle Lernen kann als Basistechnologie der KI-Entwicklungen betrachtet werden, bei welchem ein KI System auf Grundlage von Beispielen Wissen erlernt, das dann auf ein von dem KI System zu lösenden Problem angewandt werden soll. Bei dem sogenannten überwachten Lernen werden bereits klassifizierte Datensätze vorgegeben und den jeweiligen Klassen (auch „Label“ genannt) werden Eigenschaften („Features“) zugewiesen. Eine besondere Form stellt auch das aktive Lernen dar, bei dem im Verfahren dynamische Label für bestimmte Daten erfragt werden, um auf dieser Grundlage Modelle aufzubauen. Beim unüberwachten Lernen lernt das System selbstständig eine Strukturierung der Daten durch sogenannte „Cluster“. Die eingehenden Daten werden auf Grundlage eines statistischen Modells gruppiert. Das System versucht iterativ selbstständig charakteristische Muster zu entwickeln, um eine Kategorisierung vorzunehmen.

Die Auswahl des Verfahrens zum maschinellen Lernen ist zudem von den verfügbaren Daten sowie der Komplexität der Klassifikationen abhängig. Bei dem überwachten Lernen kommen Entscheidungsbäume („Decision Trees“) zum Einsatz, welche die Klassifizierungsentscheidung lenken.

Eine wesentliche Stimulierung haben KI-Entwicklungen durch das sogenannte Deep Learning erfahren, bei denen durch den Einsatz neuronaler Netze eine Leistungssteigerung des maschinellen Lernens erfolgt. Die beim Deep Learning verwendete neuronalen Netze werden mit nach bestimmten Parametern klassifizierten Daten trainiert. Die Sammlung und Aufbereitung entsprechender Trainingsdaten ist maßgeblich für den Erfolg des maschinellen Lernens. Dabei geht es um die quantitative Datenmenge sowie die Qualität der Daten. Inkompatibilitäten der Daten müssen vermieden werden und es trägt zur Beschleunigung des Lernprozesses bei, wenn bei Daten mit vielen Merkmalen eine Reduzierung der Dimensionalität auf die Merkmale stattfindet, welche für die beabsichtigte Nutzung des Lernprozesses von Bedeutung sind.

Mit den zuvor beschriebenen Maßnahmen des maschinellen Lernens, von der Auswahl der Daten, deren Aufbereitung, der Entwicklung von Decision Trees und neuronalen Netzen sowie dem Training entsprechender KI-Systeme, sind erhebliche Leistungen verbunden (zu den Einzelheiten der Methoden des maschinellen Lernens siehe *Alpaydin, Maschinelles Lernen, 2. Auf. 2019, S. 12 ff. und 23 ff.*), sodass sich die Frage nach einem geeigneten rechtlichen Schutz stellt.

I. Patentrechtlicher Schutz

Für Erfindungen im Bereich von KI-Entwicklungen gelten die Anforderungen nach § 1 PatG und Art. 52 EPÜ, d.h. diese müssen dem Gebiet der Technik zuzuordnen sein und es gilt der Patentierungsausschluss für „mathematische Methoden“ nach § 1 Abs. 3 Nr. 1 PatG/Art. 52

¹ Rechtsanwalt, Lehrbeauftragter und Honorarprofessor an der Hochschule der Medien in Stuttgart

Abs. 2 lit. a) EPÜ sowie für „Programme für Datenverarbeitungsanlagen“ nach § 1 Abs. 3 Nr. 3 PatG/Art. 52 Abs. 2 lit. c) EPÜ. Für den Nachweis einer *technischen* Lehre ist die bloße Verarbeitung und Speicherung von Daten mittels eines Computers nicht hinreichend, insbesondere wenn dies zu nicht technischen Zwecken erfolgt (*BGH, GRUR 2011, 125 – Wiedergabe topographischer Informationen; EPA, 21.11.2014 – T 1358/09*). Insoweit dürften Verfahren, welche allein die Auswertung von Daten und die Bereitstellung von Informationen betreffen, zumeist an den Patentierungsausschlüssen scheitern (*BPatG 50, 132 – Expertensystem*). Etwas anderes gilt insbesondere dann, wenn ein Verfahren zur Erzeugung und Auswertung eines Trainingsdatensatzes zu einem technischen Zweck, d.h. z.B. zur Steuerung oder Überwachung von Maschinen mittels einer Datenauswertung eingesetzt werden (*EPA, 19.04.2018 – T 2079/10*). Unabhängig von den Patentierungsausschlüssen sind KI Entwicklungen im Bereich des maschinellen Lernens auch mit den weiteren Patentierungsvoraussetzungen, d.h. insbesondere der Neuheit und erfinderischen Tätigkeit nach §§ 3, 4 PatG, Art. 54, 56 EPÜ sowie mit den Anforderungen der hinreichenden Offenbarung nach § 21 Abs. 1 Nr. 2, 34 Abs. 4, Art. 83 EPÜ konfrontiert. Es müssen die Trainingsdaten dargelegt werden (*EPA, T 0161/18*) und bei neuronalen Netzen muss erläutert werden, wie die Topologie des Netzes ausgestaltet ist und welche Gewichtung vorgenommen wurde (*BPatG 1.8.2019 – 17 W (pat) 28/19*). Denkbar ist nach den vorgenannten Grundsätzen ein patentrechtlicher Schutz für Trainings- und Analyseverfahren im Bereich des maschinellen Lernens, jedoch zumeist kein patentrechtlicher Schutz isoliert für die Daten.

II. Urheberrechtlicher Schutz nach § 69a UrhG

Im Urheberrecht hat der Schutz für Computerprogramm eine besondere Verankerung in § 69a Abs. 1 UrhG erfahren und der Schutz erstreckt sich auf ein Computerprogramm in all seinen Ausdrucksformen einschließlich des Entwurfsmaterials (*EuGH, GRUR 2011, 2020; KG Berlin, CR 2010, 424, 425*). Für den Urheberrechtsschutz irrelevant ist, in welcher Programmiersprache das Programm erschaffen wurde und ob es im Quellcode oder im ausführbaren Objektcode vorliegt (*Marly, Praxishandbuch Softwarerecht, 7. Aufl. 2018, Teil 2 Rn. 78*). Wichtig ist, dass der Begriff Computerprogramm einen Programmcode und Steuerungsfunktionen voraussetzt, das heißt reine Date oder Dateiformate ohne Steuerungsfunktionen sind vom Schutz nach § 69a UrhG nicht erfasst (*EuGH, GRUR 2012, 814 – SAS Institute*), jedoch bleibt die Möglichkeit eines Schutzes als Werk nach § 2 UrhG (*EuGH, GRUR 2011, 220 – BSA; EuGH, GRUR 2012, 814 – SAS Institute*).

Voraussetzung einen urheberrechtlichen Schutz als Computerprogramm nach § 69a UrhG ist eine hinreichende Schöpfungshöhe (*EuGH, NJW 2014, 761, 762 – Nintendo/PC Box; BGH, NJW 2000, 3212, 3213*). Bei komplexeren Computerprogrammen geht der BGH von einer tatsächlichen Vermutung für eine hinreichende Individualität der Programmgestaltung aus (*BGH, GRUR 2013, 509, 510*). Bei Methoden des Maschinenlernens ist es indes erforderlich im Einzelfall zu bestimmen, ob eine hinreichende Schöpfungshöhe besteht (*Apel/Kaulartz, Rechtlicher Schutz von Machine Learning-Modellen, Recht Digital, 2020, S. 24, 27*).

II. Schutz als Datenbank

Darüber hinaus kommt ein Schutz als Datenbank nach § 87a Abs. 1.1 UrhG in Frage. Grundsätzlich ist ein zur Schaffung des Datenbankwerkes oder zur Ermöglichung des Zuganges hierzu verwendetes Computerprogramm nicht Bestandteil des Datenbankwerkes, sodass beide Leistungen einen unterschiedlichen Schutzgegenstand begründen (*Marly, Praxishandbuch Softwarerecht, 7. Aufl. 2018, Teil 2 Rn. 102*).

Zu berücksichtigen ist, dass einzelne Elemente eines neuronalen Netzes im Einzelfall einem selbstständigen urheberrechtlichen Schutz unterliegen können. Dies gilt beispielsweise für die Gewichtungseinstellung der Verbindung zwischen den einzelnen Neuronen bei neuronalen Netzen (*Ehinger/Stiemerling, Die urheberrechtliche Schutzfähigkeit von Künstlicher Intelligenz am Beispiel von Neuronalen Netzen – Welche Strukturelemente und welche Entwicklungsphasen sind urheberrechtlich geschützt?, CR 2018, S. 761 ff.*), jedoch wird zutreffend darauf verwiesen, dass sich hierfür nur ein enger Anwendungsbereich ergibt (*Ebers/Heinze/Krügel/Steinrötter/Heinze/Wendorf, Künstliche Intelligenz und Robotik – Rechtshandbuch, 1. Aufl. 2020, § 9 Rn. 54*). Zu beachten ist bei dem urheberrechtlichen Schutz, dass jeweils nur die konkrete Ausdrucksform und – anders als beim Patentrecht – nicht die technische Lösung geschützt ist. Der urheberrechtliche Schutz von Datenbanken bezieht sich auf das Datenbankwerk nach § 4 Abs. 2 UrhG und das Leistungsschutzrecht an der Datenbank nach § 87a Abs. 1 UrhG. Während der urheberrechtliche Schutz der Datenbank an das urheberrechtliche Prinzip der Schöpfungshöhe anknüpft, bezieht sich das Leistungsschutzrecht auf die Investition in die Erstellung des Datenbankwerkes. Der Schutz bezieht sich nicht auf die im Datenbankwerk enthaltenen Einzeldaten, sondern auf das Datenbankwerk in seiner Gesamtheit.

Für den urheberrechtlichen Schutz als Datenbankwerk nach § 4 Abs. 2 UrhG ist maßgeblich, ob in der Auswahl und Strukturierung eine hinreichende geistige Schöpfung liegt (*EuGH, GRUR 2012, 386 – Football Dataco*). Zu der weiteren Voraussetzung des urheberrechtlichen Schutzes für Datenbankwerk nach § 4 Abs. 2 UrhG, nämlich der Unabhängigkeit der Elemente, wird in Bezug auf Trainingsdaten (z. B. Verkehrsdaten, Profiling-Daten etc.) darauf abgestellt, ob diese einen informatorischen Eigenwert haben (*Hacker, Immaterialgüterrechtlicher Schutz von KI-Trainingsdaten, GRUR 2020, 1025, 1029*). In Bezug auf den leistungsrechtlichen Schutz nach §§ 87a ff. UrhG ist vor allem maßgeblich, inwieweit die Datenaufbereitung mit einem Aufwand verbunden ist. Dies gilt insbesondere für die Kosten der Sammlung und Strukturierung (*EuGH, GRUR 2005, 244 – British Horseracing Board; Hacker, GRUR 2020, 1025, 1030*).

III. Wettbewerbsrechtlicher Schutz und Know-How Schutz

Die vorangehende Darstellung zeigt die Grenzen des patentrechtlichen und urheberrechtlichen Schutzes für KI Trainingsdaten.

Der wettbewerbsrechtliche Schutz nach § 4 Nr. 3 UWG schließt diese Lücke nur in Ausnahmefällen, d.h. wenn weitere Umstände hinzutreten, welche die Unlauterkeit begründen. Dies gilt beispielsweise dann, wenn im Rahmen einer Kooperation nur für diesen spezifischen

Zweck zugänglich gemachte Trainingsdaten unberechtigt genutzt werden und insoweit die in § 4 Nr. 3 (c) UWG geregelte Fallgruppe greift. Auch in diesen Fällen ist es jedoch erforderlich, den Nachweis zu führen, dass die Trainingsdaten eine wettbewerbliche Eigenart aufweisen. Dies ist dann der Fall, wenn die Trainingsdaten Elemente aufweisen, die eine herkunftshinweisende Funktion haben (*BGH, Urteil v. 16.11.2017, IZR 91/16, Tz. 14*; *BGH, Urteil. V. 14.9.2017; IZR 2/16, Tz. 20*). Hierin dürfte eine wesentliche Hürde liegen, da die Individualisierung und abnehmerseitige Zuordnung von Trainingsdaten häufig nur schwer zu begründen sein dürfte.

Darüber hinaus kommt bei KI Trainingsdaten ein Schutz als Geschäftsgeheimnis in Frage. Diesbezüglich ist eine Stärkung durch die Einführung des Geschäftsgeheimnisgesetzes im Jahr 2019 eingetreten, durch welches die EU-Richtlinie 2016/943 umgesetzt wurde. Voraussetzung für einen Schutz als Geschäftsgeheimnis sind insbesondere auch angemessene Geheimhaltungsmaßnahmen nach § 2 Abs. 1 lit. b) GeschGehG). Relevant ist mithin, dass in der Entwicklung und Anwendung der Maschinenlernkonzepte und der Trainingsdaten angemessene Geheimhaltungsmaßnahmen getroffen werden. Dies gilt insbesondere für die Beschränkung des Zugangs zu den entsprechenden Systemen. Weiterhin ist auch eine Auferlegung von Geheimhaltungsverpflichtungen maßgeblich. Diesbezüglich ist zu erwarten, dass im Bereich der KI-Entwicklungen unterschiedliche Ansätze zur Anwendung gelangen werden, nämlich ein auf die allgemeine Nutzung von Trainingsdaten ausgerichteter Ansatz, der auf eine weite Verbreitung und Bündelung von Trainingsdaten zielt und auf der anderen Seite ein durch den Datenhandel geprägter Ansatz, der darauf zielt, durch die Geheimhaltung Lizenzgeschäftsmodelle abzusichern. Das Risiko des Schutzes als Geschäftsgeheimnis liegt darin, dass durch eine unintendierte oder auch unberechtigte Preisgabe der Schutz verloren geht und nur mögliche Schadensersatzansprüche gegenüber dem Preisgebenden bleiben. Unabhängig von dem zuvor beschriebenen Risiko des Verlustes der Geheimhaltung ist mit dieser rechtlichen Schutzform auch das Problem verbunden, dass sich der Verzicht auf die Preisgabe innovationshemmend auswirkt. Gerade mit Blick auf die einleitend beschriebene Bedeutung von Trainingsdaten für die KI Entwicklung ist eine Offenlegung grundsätzlich wünschenswert. Gleichzeitig müssen auch wirtschaftliche Anreizmodell für die Aufbereitung von Trainingsdaten geschaffen werden, die nur darauf basieren können, dass ein adäquater rechtlicher Schutz für Trainingsdaten geschaffen wird, der eine Generierung von Lizenzeinnahmen für die Bereitstellung erlaubt.