CENTRE DE DROIT
BANCAIRE ET FINANCIER

Konzepte zu unterscheiden

KI-Systeme und KI-Modelle mit allgemeinem Verwendungszweck

Par Yannick Caballero Cuevas le 4 November 2024

Seit dem 1. August 2024 ist die europäische Verordnung über künstliche Intelligenz (KI-VO) in Kraft (vgl. cdbf.ch/1359/). Sie gilt sowohl für Systeme künstlicher Intelligenz (KI-Systeme) als auch für allgemein verwendbare KI-Modelle (Art. 2 KI-VO). Dieser Kommentar konzentriert sich auf die Definitionen dieser beiden Schlüsselbegriffe bei der Anwendung der KI-VO und versucht, ihre Merkmale und Besonderheiten hervorzuheben.

A. Der Begriff der SIA

Gemäss Art. 3 Ziff. 1 KI-VO ist ein Ki-System „ein **maschinengestütztes System**, das für einen in **unterschiedlichem Grade autonomen Betrieb** ausgelegt ist und das nach seiner Betriebsaufnahme **anpassungsfähig** sein kann und das aus den **erhaltenen Eingaben** für explizite oder implizite Ziele ableitet, wie Ausgaben wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen **erstellt werden**, die physische oder virtuelle Umgebungen beeinflussen können“. Die verschiedenen hervorgehobenen Begriffe bedürfen einiger Erläuterungen.

Zunächst einmal wird zur Definition eines KI-Systems in der französischen Version die Formulierung „système automatisé“ verwendet, während in der englischen und deutschen Version die Begriffe „*machine-based system*“ und „*maschinengestütztes System*“ verwendet werden. Diese Formulierungen besagen lediglich, dass ein KI-System ein System ist, das mithilfe einer Maschine wie einem Computer, einem selbstfahrenden Auto oder einem Mobiltelefon funktioniert.

Zweitens muss das KI-System aus den Eingaben (*Input*) ableiten, wie es die Ausgaben (*Output*) erzeugt, die die Form einer Vorhersage, eines Inhalts (z. B. Text, Bild, Video usw.), einer Empfehlung oder auch einer Entscheidung annehmen können. Die Liste der Beispiele ist nicht erschöpfend. Die vom KI-System erzeugte Ausgabe muss die physische oder virtuelle Umgebung beeinflussen. Unserer Ansicht nach muss diese Präzisierung im Zusammenhang mit den Zielen der KI-VO gelesen werden, die insbesondere darauf abzielen, ein hohes Maß an Schutz der Gesundheit, der Sicherheit und der Grundrechte zu gewährleisten (Erw. 1 KI-VO). Viele deterministische Algorithmen erzeugen auch Ausgaben aus Eingaben, so dass dieses Merkmal ein KI-System nicht von einem *if-then*-System unterscheidet.

Das KI-System muss einen gewissen Grad an Autonomie aufweisen, was bedeutet, dass es in

seinem Handeln Unabhängigkeit genießt (Erw. 12 KI-VO). Die KI-VO legt jedoch nicht fest, welchen Grad an Autonomie ein System aufweisen muss, um als KI-System eingestuft zu werden. Unserer Ansicht nach weisen Systeme, die auf der Grundlage eines *Machine-Learning-Ansatzes* entwickelt wurden, einen ausreichenden Grad an Autonomie auf. Dieses Kriterium der Autonomie eines KI-Systems ermöglicht es, ein KI-System von einer gewöhnlichen Computeranwendung zu unterscheiden. So sind beispielsweise Systeme, die auf einer rein deterministischen Logik basieren, von der Definition eines KI-System ausgeschlossen. Es ist hilfreich, darauf hinzuweisen, dass eine menschliche Überwachung eines KI-Systems nicht zu dem Schluss führt, dass das System keine Autonomie besitzt, da die KI-VO für KI-Systeme mit hohem Risiko ausdrücklich eine menschliche Kontrolle verlangt (vgl. Art. 14 KI-VO).

Schliesslich enthält die Definition ein letztes Merkmal, indem sie klarstellt, dass AIS nach ihrer Einführung angepasst werden können. Die Verwendung des Verbs „können“ zeigt, dass dieses Merkmal im Gegensatz zum Kriterium des Autonomiegrads nicht wesentlich ist, um ein System als AIS zu qualifizieren.

B. Der Begriff des KI-Modelld mit allgemeinem Verwendungszweck

Nach Art. 3 Ziff. 63 KI-VO ist ein KI-Modell mit allgemeinem Verwendungszweck „ein KI-Modell — einschliesslich der Fälle, in denen ein solches KI-Modell mit einer **großen Datenmenge unter umfassender Selbstüberwachung trainiert wird** —, das eine **erhebliche allgemeine Verwendbarkeit** aufweist und in der Lage ist, unabhängig von der Art und Weise seines Inverkehrbringens ein **breites Spektrum unterschiedlicher Aufgaben** kompetent zu erfüllen, und das **in eine Vielzahl nachgelagerter Systeme oder Anwendungen** integriert werden kann, ausgenommen KI-Modelle, die vor ihrem Inverkehrbringen für Forschungs- und Entwicklungstätigkeiten oder die Konzipierung von Prototypen eingesetzt werden“ (Hervorhebung hinzugefügt). Dieses Konzept wurde während der Diskussionen im Europäischen Parlament eingeführt und entspricht dem Wunsch, KIs wie ChatGPT oder Gemini zu regulieren. Seine Hinzufügung wirft jedoch Fragen auf. Ist ein KI-Modell mit allgemeinem Verwendungszweck zwangsläufig ein KI-System ? Schliessen sich die Definitionen von KI-System und KI-Modell mit allgemeinem Verwendungszweck gegenseitig aus ?

Ein KI-Modell mit allgemeinem Verwendungszweck weist vier Hauptmerkmale auf. Erstens muss das Training dieser Modelle mit großen und vielfältigen Datensätzen erfolgen und sich auf Selbstüberwachung stützen (z. B. *Deep Learning*). Zweitens darf das Modell nicht für eine bestimmte Aufgabe konzipiert sein, sondern muss in der Lage sein, eine breite Palette unterschiedlicher Aufgaben zu erfüllen, wie z. B. die Übersetzung von Texten, die Generierung von Text-, Bild- oder audiovisuellen Inhalten oder die Analyse von Dokumenten. Die Allgemeingültigkeit eines Modells kann auch durch die Anzahl der Parameter bestimmt werden, die das Modell haben kann (Erw. 98 KI-VO). Schließlich muss ein KI-Modell in verschiedene Systeme (z. B. Smartphones, vernetzte Objekte usw.) oder Anwendungen integriert werden können.

Der Begriff der KI-Modell mit allgemeinem Verwendungszweck ist vom Begriff des KI-Systems zu unterscheiden (Erw. 97 KI-VO). Ein allgemein verwendbares KI-Modell ist nicht per se ein KI-System, kann aber eine wesentliche Komponente eines AIS darstellen. In diesem Fall spricht man von einem KI-System mit allgemeinem Verwendungszweck, wenn dieses System in der Lage ist, verschiedene Verwendungszwecke zu erfüllen (Erw. 100 KI-VO). Nehmen wir das

Beispiel ChatGPT : ChatGPT verwendet ein *Large Language Model* (LLM) zur Generierung von Inhalten. Das verwendete LLM ist ein KI-Modell mit allgemeinem Verwendungszweck im Sinne der KI-VO. OpenAI hat dann eine Schnittstelle entwickelt, die es den Nutzern ermöglicht, mit dem LLM zu interagieren. Folglich stellt der LLM eine wesentliche Komponente der Schnittstelle dar, die unserer Meinung nach als KI-System bezeichnet werden kann.

Schlussfolgerung

Die Hardware-Anwendung von KI-VO beruht im Wesentlichen auf den Begriffen KI-System und KI-Modell mit allgemeinem Verwendungszweck. Während ein KI-System ein maschinengestütztes System mit einem gewissen Grad an Autonomie ist, das in der Lage ist, Ausgaben zu generieren, die physische oder virtuelle Umgebungen beeinflussen, konzentriert sich ein KI-Modell mit allgemeinem Verwendungszweck auf die Fähigkeit einer KI, eine breite Palette von Aufgaben auszuführen. Wichtig zu beachten ist, dass ein KI-Modell an sich kein KI-System ist, aber ein wesentlicher Bestandteil davon sein kann. Die Integration von KI-Modellen erfolgt in der Regel über Benutzerschnittstellen oder APIs, wie z. B. OpenAI mit ChatGPT oder dessen API. Das KI-Büro – das für die Umsetzung von KI-VO auf europäischer Ebene zuständig ist – wird den Unternehmen wahrscheinlich dabei helfen müssen, zwischen diesen beiden grundlegenden Konzepten zu navigieren.
