Büro für Technikfolgen-Abschätzung
beim Deutschen Bundestag

TAB

Themenkurzprofil Nr. 78
Februar 2025

Large Action Models – neue KI-basierte Interaktionsformen mit digitalen Technologien

Mona Hille

In Kürze

In jüngster Zeit hat ein neuer Trend an Dynamik gewonnen, bei dem generative KI-Agenten in die Lage versetzt werden, eigenständig Aufgaben für die Nutzer/innen auszuführen. Dieses Themenkurzprofil fokussiert sich auf große Aktionsmodelle (Large Action Models – LAM), die eine sogenannte neurosymbolische Modellarchitektur aufweisen. Dabei handelt es sich um die Kombination von symbolischer und neuronaler KI, die eine neuartige Steuerung digitaler Anwendungen ermöglicht. Technisch vermitteln LAM zwischen menschlicher Intention und digitaler Ausführung, in etwa wie Betriebssysteme zwischen Hardware und Software. Neurosymbolische KI-Modelle werden in LAM-basierten Agenten eingesetzt. Gegenwärtig sind erst wenige LAM-basierte Agenten als Apps oder eigenständige Geräte auf dem Markt verfügbar. Sie zeichnen sich durch eine vereinfachte Interaktionsform für den menschlichen Nutzer aus und sollen diverse Handlungsketten im digitalen Raum autonom ausführen können. Perspektivisch ist davon auszugehen, dass Menschen zukünftig zur Erledigung diverser Aufgaben ihren mobilen Endgeräten Sprachbefehle erteilen und dann die integrierten KI-Agenten selbst scrollen, tippen und klicken.

Offen ist, wie Datenschutz und der Schutz der Privatsphäre gewährleistet werden können, wenn KI-Agenten als Alltagsassistenten durch die Ausführung ihrer Aufgaben umfassende Datenprofile der Nutzer/innen erstellen.

Angesichts der Geschwindigkeit, mit der KI-Systeme weiterentwickelt werden, kann davon ausgegangen werden, dass die gegenwärtig bekannten KI-Agenten den Anfang einer veränderten menschlichen Interaktion mit digitalen Technologien markieren. Die dynamische Entwicklung autonomer KI-Agenten macht es schwer, die Folgewirkungen auf die individuelle Problemlösungskompetenz sowie auf die persönliche Entscheidungsfreiheit des Menschen abzuschätzen.

Hintergrund und Entwicklungsstand

Die Informationstechnologie wurde in den letzten Jahren geprägt von enormen Fortschritten in der Weiterentwicklung von KI. Insbesondere die Erfolge des Deep Learnings haben diese Entwicklungen hervorgerufen und diverse Anwendungs- und Einsatzmöglichkeiten von KI ermöglicht. Deep Learning ist ein Teilbereich des maschinellen Lernens, der auf künstlichen neuronalen Netzen basiert. Die tiefen Netze mit vielen Schichten sind besonders gut geeignet, komplexe Muster in großen Datenmengen zu erkennen. Deep Learning gilt im Bereich der KI als wichtiger Entwicklungsschritt und führte in verschiedenen Bereichen der Informatik, wie z. B. bei der Spracherkennung, dem Sehen und Verstehen von Bildern, der Verarbeitung natürlicher Sprache und der maschinellen Übersetzung, zu enormen Fortschritten (d'Avila Garcez/Lamb 2023). Im November 2022 hat Open AI mit ChatGPT den Zugang zu KI für jedermann ermöglicht und damit einhergehend unzählige Weiterentwicklungen generativer KI-Modelle auf Basis von großen Sprachmodellen (Large Language Models – LLM) angestoßen. In der Folge wurden seit Sommer 2023 diverse KI-Agentensysteme veröffentlicht, die mit verschiedenen Ansätzen ein gemeinsames Ziel verfolgen: die autonome Durchführung von Nutzeraufgaben. Zu diesen zählen der AI-Pin (LLM-Agentengerät, November 2023), das R1 (LAM-Gerät, Januar 2024), das T-Phone AI-Concept (LAM-App, Februar 2024), Apples ReALM (~LAM, März 2023), Salesforces xLAM (LAM, September 2024), Antropics Claude 3.5 Sonnet und Claude 3.5 Haiku (LLM-Agent, Oktober 2024), Microsofts OmniParser (LLM-Agent, Oktober 2024), Gemini 2.0 Google (LLM-Agent, Dezember 2024), Galaxy S25 (Galaxy AI/Gemini 2.0, LLM-Agent, Januar 2025) u. v. m.

Perspektivisch kann davon ausgegangen werden, dass autonome KI-Agenten, gleich welcher Modellstruktur, die Nutzung digitaler Plattformen für den Menschen effizienter machen (Singhal/Singh 2024). Wegen der für generative KI-Modelle neuartigen neurosymbolischen Programmarchitektur fokussiert dieses Themenkurzprofil auf die LAM-basierten Agenten.

Begriffsklärung und Abgrenzung

Um die Begrifflichkeiten der nachfolgenden Ausführungen besser zu verstehen und einzuordnen, erfolgt an dieser Stelle zunächst eine Differenzierung zwischen den zentralen Begriffen KI-Systeme, KI-Modelle und KI-Agenten.

- KI-Modelle sind die Grundlage für jede Form von künstlicher Intelligenz. Sie sind mathematische oder algorithmische Strukturen, die mit Daten trainiert werden, um spezifische Aufgaben zu lösen. Am bekanntesten sind die großen Sprachmodelle (Large Language Models – LLM), die speziell auf die Verarbeitung und Generierung natürlicher Sprache ausgelegt sind.
- KI-Agenten sind Anwendungen oder Instanzen, die auf KI-Modellen basieren und Aufgaben in der realen oder digitalen Welt ausführen. Der Begriff KI-Agent ist allgemein und umfasst alle Systeme, die KI nutzen, unabhängig davon, auf welchem Modell sie basieren. Dabei kann ein KI-Agent einfache Automatisierungen ausführen oder komplexe Entscheidungen treffen.
- Autonome KI-Agenten stellen eine Unterkategorie der KI-Agenten dar. Sie sind in der Lage, ohne menschliche Eingriffe Entscheidungen zu treffen und Aktionen auszuführen. Autonomie wird dabei durch das zugrunde liegende KI-Modell und die spezifischen Systemstrukturen erreicht.
- Ein KI-System ist das Gesamtkonstrukt, das in einem spezifischen Anwendungsbereich eingesetzt wird. Es umfasst nicht nur den KI-Agenten und das zugrunde liegende Modell, sondern auch alle erforderlichen Schnittstellen, Datenquellen und die technische Infrastruktur. Während der KI-Agent operative Aufgaben übernimmt, bietet das KI-System den Rahmen für die Implementierung und Nutzung der Technologie.

Was sind LAM?

Große Aktionsmodelle (Large Action Models, LAM) sind eine neuartige Form generativer künstlicher Intelligenz. LAM verbinden die klassische regelbasierte symbolische KI mit der auf Mustererkennung ausgerichteten neuronalen KI (Collins 2024). Diese hybride Form der neurosymbolischen KI befähigt LAM dazu, Muster in Aktionen zu erkennen und darauf aufbauend Handlungsabfolgen zu planen und auszuführen, um eine vorgegebene Aufgabe autonom zu erledigen (Kautz 2022). Das Training der LAM erfolgt in einer digitalen Umgebung, in der reale Benutzeroberflächen simuliert und das menschliche Nutzerverhalten bildbasiert demonstriert werden (Singhal/Singh 2024). LAM erkennen die Ausführungslogik und werden in die Lage versetzt, menschliche Absichten zu identifizieren, um sie in eigene umsetzbare Aufgaben zu übersetzen. Praktisch sollen LAM genau wie menschliche Nutzer/innen mit jedweder Anwendungsoberfläche über Klicks, Scrollen und das Ausfüllen von Textfeldern interagieren können (d'Avila Garcez/Lamb 2023; Fauscette o. J.). Mit LAM-basierten Agenten wird das Ziel verfolgt, den seit der Einführung des iPhones im Jahr 2007 eingeübten Umgang mit App-basierten Smartphones, Tablets und Smartwatches durch die autonome Systemsteuerung des neurosymbolischen KI-Modells abzulösen. Die Nutzer/innen sollen von der Bildschirmabhängigkeit befreit und das Smartphone durch die veränderte Steuerung in seiner etablierten Funktionsweise auf lange Sicht ersetzt werden. Diese neue Art und Weise der Interaktion zwischen Mensch und digitaler Technologie soll dadurch erreicht werden, dass der LAM-basierte Agent die menschliche Vorgehensweise bei der Bedienung von Menüs und Schaltflächen sowie die Abfolge von Bedienschritten nachahmt.

Die spezifische Modellstruktur ermöglicht es den LAM-basierten Agenten, die Logik hinter den Prozessschritten bei der Bedienung einer Benutzeroberfläche zu erkennen und damit einerseits fremde Webseiten und Apps, auf die sie nicht trainiert wurden, zu bedienen, und andererseits eigenständig die notwendigen Schritte zur Lösung einer Aufgabe zu entwickeln (Savarese 2024). Die Steuerung soll über sprachbasierte Anweisungen erfolgen; der KI-Agent führt die Aufgabe dann einschließlich aller erforderlichen Zwischenschritte selbstständig aus, z. B. die Gestaltung eines Reiseplans inklusive der Auswahl passender Flugverbindungen und Hotels. Im Unterschied zu LLM-Agenten, die als spezialisierte Erweiterung eines großen Sprachmodells den Zugriff auf Anwendungsprogrammierschnittstellen (Application Programming Interfaces – API) benötigen, um spezifische Aufgaben erledigen oder bestimmte Aktionen ausführen zu können, können LAM-basierte Agenten Benutzeroberflächen bedienen, ohne dabei auf APIs angewiesen zu sein.

Erste LAM-basierte Produkte erreichen Marktreife

Seit 2024 werden LAM in eigenständige Geräte integriert oder als App z. B. für Smartphones oder Tablets entwickelt. Ein Beispiel für ein eigenständiges Gerät, das auf einem LAM (rabbit inc. o. J.a) basiert, ist das vom Unternehmen rabbit inc. entwickelte R1, das im Januar 2024 auf der Technikmesse CES 2024 in Las Vegas vorgestellt wurde. Das per Sprachbefehl steuerbare R1 soll nach Unternehmensangaben in der Lage sein, komplette Handlungsketten im digitalen Raum auszuführen, und damit eine Art universeller Controller für Apps sein (Eliaçık 2024). Im Trainingsmodus können die Nutzer/innen das R1 durch Demonstration ihres Nutzerverhaltens selbstständig trainieren (rabbit inc. o. J.c). Die neurosymbolische Modellarchitektur ermöglicht nach Angaben des Unternehmens die Verarbeitung der sprachbasierten Handlungsaufforderungen ausschließlich im lokalen Gerät. Dadurch soll einerseits die Privatsphäre der Nutzer/innen gewahrt bleiben und andererseits der Energieverbrauch von LAM im Vergleich zu LLM geringer sein. Zur Nutzung des Geräts wird für jede/n Käufer/in ein Account auf der von rabbit inc. betriebenen Plattform rabbithole eingerichtet. Damit das Gerät eigenständig für den/die Nutzer/in mit Anwendungen von Drittanbietern interagieren kann, kann der/die Nutzer/in ihre Zugangsdaten zur jeweiligen Authentifizierung auf rabbithole speichern.

Dabei ist unklar, wie genau das R1 auf die Zugangsdaten der jeweiligen Drittanbieter zugreift und Zahlungsprozesse ausgeführt werden. Umfassende technische Details oder eine wissenschaftliche Beschreibung der Technologie liefert rabbit inc. nicht. Hinsichtlich des Datenschutzes gibt das Unternehmen an, die persönlichen Zugangsdaten nicht selbst zu verwalten oder zu speichern. Trotzdem hätte rabbit inc. aber möglicherweise über das Modell und seine Dienste Zugriff auf laufende Websessions und könnte damit in der Lage sein, diese auszuwerten (Floemer 2024; Grüner 2024; rabbit inc. 2024; Weiß 2024; Weiss 2024). Erste Rezensionen des Gerätes reichen von Begeisterung bis zu Ernüchterung. Es zeigt sich einerseits, dass das R1 das Funktionsversprechen grundsätzlich erfüllt, und andererseits, dass es sich um eine Technologie handelt, die noch auf technische Optimierung und Weiterentwicklung angewiesen ist.[1] In Reaktion darauf hat das Start-up zahlreiche Updates umgesetzt (rabbit inc. o. J.b).

Hypothetisches Anwendungsszenario für LAM

„Ein/e Versicherungsmakler/in trifft sich mit einem Kunden in einem Online-Meeting und bespricht dessen Bedürfnisse, Ideen und mögliche nächste Schritte. Ein Transkript des Gesprächs wird automatisch aufgezeichnet und mit weiteren relevanten Kunden- und Prozessinformationen organisiert. Nach dem Gespräch überprüft das LAM das Transkript, fasst die wichtigsten Momente zusammen und sendet das Ergebnis an den Versicherungsvertreter, damit er es zu einem späteren Zeitpunkt einfach überprüfen kann. Darüber hinaus identifiziert das LAM die nächsten Schritte, die unternommen werden sollten, wie z. B. die Bereitstellung zusätzlicher Informationen, die in dem Gespräch erwähnt wurden, für den Kunden. Auf Grundlage dieser Erkenntnisse wird automatisch eine Follow-up-E-Mail verfasst, gefolgt von einer Suche in der Unternehmensliteratur nach allen relevanten Dokumenten, die als Anhang beigefügt sein können. Der Versicherungsvertreter wird dann benachrichtigt, dass die nächsten Schritte ausgeführt werden können, wobei eine letzte Bestätigung und ein kurzes Korrekturlesen vor der Ausführung möglich sind. Da das LAM die Arbeitsvorgänge der Versicherungsagentur kennt, kann es weitere Schritte vorschlagen, um den Versicherungsvertreter produktiv und fokussiert zu halten, sei es ein Folgegespräch mit einer automatisch vorgeschlagenen Agenda oder eine Upsellingmöglichkeit auf der Grundlage einer früheren Kundenentscheidung. Auf dem Weg dorthin kann das LAM auf Anzeichen achten, die darauf hindeuten, dass andere Beteiligte einbezogen werden müssen. Beispielsweise kann ein Kunde, der Anzeichen von Frustration oder Zögern zeigt, als ‚gefährdete Kundenbeziehung' eingestuft und an einen Kundendienstspezialisten verwiesen werden, der sich speziell auf die Erhaltung der Zufriedenheit konzentriert." [Übersetzung durch die Autorin] (Savarese 2024)

Weitere Entwickler von LAM-basierten Produkten sind (Stand September 2024) ein Unternehmensverbund aus Deutscher Telekom, Qualcomm Technologies Inc. und Brain.ai, der ein digitales Assistenzsystem für das Smartphone auf den Markt bringen möchte, sowie das Softwareunternehmen Salesforce,

1 Die ersten 100.000 Exemplare des R1 wurden im April 2024 ausgeliefert. Seit August 2024 kann das Gerät weltweit erworben werden.

das mit dem xLAM-Modell ein neues KI-System für den digitalen Dienstleistungssektor entwickelt hat.

Die Deutsche Telekom und ihre Mitstreiter setzen – anders als rabbit inc. – mit ihrer Innovation zwar darauf, die vielfältigen Apps des klassischen Smartphones zu ersetzen, das Smartphone als solches, das auch weiterhin wie gewohnt bedient werden kann, jedoch als Hardware beizubehalten. Vorgestellt wurde der Prototyp auf dem Mobile World Congress im Februar 2024, ein Zeitraum für die Markteinführung wurde nicht bekannt gegeben (Fischer 2024; Mühlroth 2024).

Das Softwareunternehmen Salesforce bietet seit September 2024 die LAM-Modellreihe xLAM an, mit der nicht nur einzelne Aufgaben, sondern ganze Prozesse für die Kund/innen des Unternehmens automatisiert werden sollen. Die unterschiedlichen Large-Action-Modelle von Salesforce sind darauf ausgelegt, die Integration von KI-Agenten in diverse Arbeitsabläufen zu vereinfachen und zu optimieren. Beispielsweise sollen im Rahmen des Kundenbeziehungssystems (Customer Relationship Management – CRM) mithilfe von LAM-basierten Agenten alle Interaktionen mit aktuellen und potenziellen Kund/innen autonom verwaltet werden (Beispiel Textkasten). Mit der xLAM-Reihe sollen unterschiedlichste Anforderungen von Benutzer/innen und Rechenkapazitäten abgedeckt werden, um den Zugang zu leistungsstarken Agentenfunktionen zu erleichtern und diese an reale Anwendungen anzupassen. So umfasst die Modellreihe von Salesforce verschiedene LAM, die sich in ihrer Größe[2] unterscheiden, da sie für verschiedene Anwendungsfälle entwickelt wurden. Sie basieren auf den LLM von Mixtral Instruct und zielen je nach Anwendungsfall auf eine ausgewogene Leistung in einem breiten Spektrum von Agentenaufgaben ab. Laut Salesforce könnten mit verschiedenen hochspezialisierten digitalen Agenten immer komplexere Aufgaben autonom erledigt werden, wie beispielsweise die Kollaboration von KI-Agenten zur Abstimmung von Terminen zweier Personen. Salesforce spricht in dem Kontext von einem Grundstein für robuste autonome Agentensysteme (Zhang et al. 2024).

2 So soll sich das kleinste Modell Tiny (xLAM-1B) für geräteinterne Anwendungen eignen. Das nächstgrößere Modell Small (xLAM-7B) soll für schnelle akademische Aufgaben mit begrenzten GPU-Ressourcen eingesetzt werden. Für industrielle Anwendungen, die höhere Anforderungen an die Latenz, den Ressourcenverbrauch und die Leistung erfordern, eignet sich das Modell Medium (xLAM-8x7B). Das größte Modell der xLAM-Reihe Large (xLAM-8x22B) dient dem Einsatz in Prozessen, die sehr hohe Rechenkapazitäten erfordern (Zhang et al. 2024).

Unklare künftige Entwicklungsrichtung für LAM

Obwohl bisher nur wenige LAM-basierte Produkte auf dem Markt verfügbar sind, deutet sich mit ihrer Verfügbarkeit ein neuer Entwicklungspfad in der Mensch-Technik-Interaktion an. Die vorgestellten LAM-basierten Agentensysteme könnten Vorboten einer zukünftigen Autonomisierung von KI-Agenten im privaten und beruflichen Alltag sein. Welchen Mehrwert KI-Agenten wie das R1 für den privaten sowie die xLAM-Entwicklungen von Salesforce für den professionellen Einsatz tatsächlich haben, lässt sich wenige Monate nach der Markteinführung dieser Systeme allerdings noch nicht beurteilen. Noch befinden sie sich in der Optimierungsphase. Die Entwickler/innen korrigieren Systemfehler mit Updates, und durch die breite Anwendung von privaten und wirtschaftlichen Nutzer/innen werden auch die Trainingseffekte die KI-Modelle weiter verbessern. Da LAM in der Lage sind, diverse Webseiten, spezifische Apps und perspektivisch auch über Smart-Home-Systeme vernetzte physische Geräte autonom zu steuern, stellen sich dennoch bereits heute Fragen nach einer verantwortungsvollen Integration dieser KI-Modelle in berufliche und private Umgebungen (Singhal/Singh 2024).

Gesellschaftliche und politische Relevanz

Chancen und Risiken, die von autonomen KI-Agenten ausgehen, sind aufgrund des frühen Entwicklungsstadiums noch relativ unklar. Hinsichtlich der gesellschaftlichen und politischen Relevanz lässt sich aber bereits festhalten, dass LAM aufgrund ihres Ressourcenverbrauchs ähnliche ökologische Auswirkungen haben wie die seit 2022 verfügbaren generativen KI-Modelle (TAB 2023). Möglicherweise können LAM hinsichtlich der Energieeffizienz von KI-Systemen aber auf lange Sicht positive Effekte mit sich bringen, da sie aufgrund der neuartigen Modellarchitektur eine verbesserte Latenzzeit haben. Statt die Berechnungen über ausgelagerte Server durchführen zu müssen, können LAM-Systeme direkt auf den mobilen Endgeräten laufen (Moniz et al. 2024; Yasir 2024).

Darüber hinaus muss davon ausgegangen werden, dass die Nachvollziehbarkeit und Vertrauenswürdigkeit der LAM-generierten Vorschläge, beispielsweise für Reisebuchungen, die

Auswirkungen auf den Datenschutz und die Datensouveränität sowie auch der Schutz der Privatsphäre zentrale Aspekte sind, von denen Risiken für die Nutzer/innen ausgehen könnten.

Vertrauenswürdigkeit

Während die Frage nach der Vertrauenswürdigkeit von KI-Systemen, beispielsweise mit Blick auf die Herstellung von Deepfakes, bisher maßgeblich gegenüber den von KI erzeugten Artefakten diskutiert wurde, muss bei der Anwendung/Nutzung von LAM-basierten Agenten nun hinterfragt werden, wie vertrauenswürdig deren Vorschläge und Handlungen sind. Um Vertrauen in die Technologie zu schaffen, bedarf es transparenter Einblicke zumindest in die Entscheidungsprozesse der KI. Während generative KI-Modelle als Blackbox gelten, wird LAM zugeschrieben, durch ihre neurosymbolische Modellstruktur Handlungsschritte nachvollziehbar zu machen. Dies liegt daran, dass der symbolische Teil des Modells regelbasiert vorgeht und so die Prozessschritte, die das LAM zur Entwicklung eines Handlungspfades durchläuft, technisch nachvollzogen werden können. Hinsichtlich der Forschungs- und Entwicklungsarbeiten im Forschungsfeld der Erklärbarkeit von KI könnte die den großen Aktionsmodellen (LAM) zugrunde liegende Modellstruktur eine relevante Weiterentwicklung von KI anstoßen (Qiu et al. 2024; Tabrez 2024).

Datenschutz und Datensouveränität

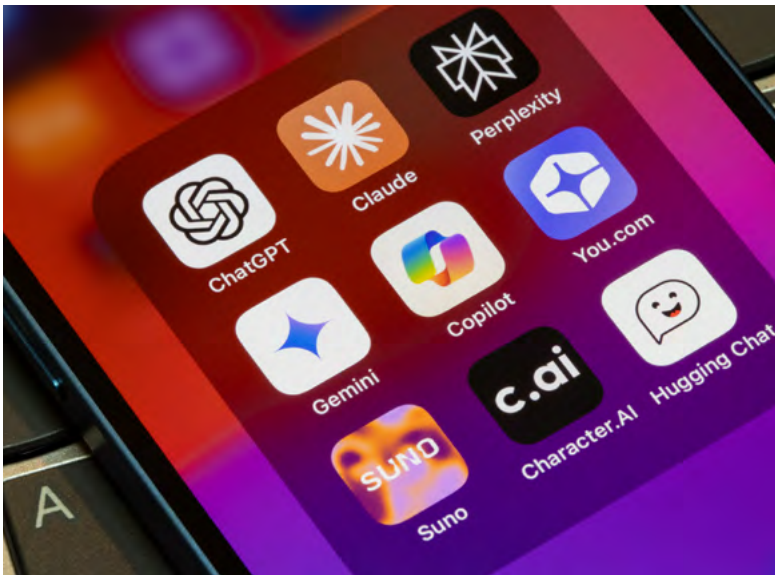
Perspektivisch ergeben sich weitere Risiken wie Missbrauch und Betrug dadurch, dass LAM-basierte Anwendungen zur Durchführung autonomer Handlungen und Entscheidungen Zugriff auf persönliche Zugangsdaten der Nutzer/innen zu Webportalen von Drittanbietern benötigen. Konkrete Auswirkungen hinsichtlich der Datensouveränität und des Datenschutzes können gegenwärtig nicht evidenzbasiert diskutiert werden. Erst mit der breiteren gesellschaftlichen Nutzung der Systeme kann eine vertiefte Auseinandersetzung erfolgen, beispielsweise entlang der Grundlagen der Datenschutz-Grundverordnung (DSGVO)[3] sowie der Verordnung über künstliche Intelligenz (KI-Verordnung)[4].

Angesichts der anvisierten Nutzung als KI-Agent, der im Auftrag der Nutzer/innen eigenständig Aufgaben ausführt und Zugang zu einer großen Menge personenbezogener Informationen hat, ist aber davon auszugehen, dass diese Technologie ein zusätzliches Risiko für den Schutz der persönlichen Daten wie auch der Privatsphäre birgt und sich dadurch das Schadenspotenzial im Falle von Missbrauch durch Dritte erhöht. Mit datenschutzrechtlichen Risiken könnte z. B. der Demonstrationsmodus der eigenen Nutzungsmuster von Apps und Webseiten im Rahmen des individuellen Trainings des eigenen LAM-basierten Agenten verbunden sein, da dieser zu einer Erweiterung des persönlichen Datensatzes (digitale Identität) um die individuellen Interaktionsdaten führt. Zum gegenwärtigen Zeitpunkt ist unklar, inwieweit diese Interaktionsdaten an die Anbieter von LAM-Anwendungen weitergegeben werden und ob eine Einflussnahme Dritter auf die Aktionen und Vorschläge des KI-Agenten möglich ist. Bereits heute werden vielfältige personenbezogene Daten verwendet, um Nutzer/innen individualisierte Angebote zu machen und ihre Entscheidungen zu beeinflussen – die Nutzung von KI-Agenten könnte diese Entwicklung noch verstärken (Diemerling et al. 2024; Redaelli 2024; Simon et al. 2024).

Darüber hinaus scheint es nicht ausgeschlossen, dass auch biometrische Daten aus den Sprachbefehlen der Nutzer/innen, also der Stimme, genutzt werden können, um die darin enthaltenen Informationen, beispielsweise zum Gesundheitszustand, in die Entscheidungsprozesse des LAM-basierten Agenten einzubeziehen. Hierbei könnte es zu algorithmischer Diskriminierung und darüber hinaus zu Einschränkungen der Entscheidungsfreiheit der Nutzer/innen durch eine prädiktive Modellierung der Entscheidungen kommen (Diemerling et al. 2024; Jedličková 2024; Prunkl 2024; Singhal/Singh 2024). Denkbar wäre etwa, dass der KI-Agent bei der Kuratierung einer Reise Sehenswürdigkeiten auf hohen Gebäuden oder Brücken ausschließt, weil aus der Stimme auf eine psychische Erkrankung wie Depression geschlossen werden kann. Das Modell könnte so programmiert sein, dass Situationen vermieden werden sollen, in denen Suizid begangen werden könnte.

Veränderung des Interaktionsmodus als Chance und Risiko

Das Innovationspotenzial von LAM besteht in der Autonomisierung von KI-Agenten. Digitale Anwendungen, die immer mehr Aufgaben für den Menschen übernehmen, bergen sowohl Chan-



3 Verordnung (EU) 2016/679 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung)

4 Verordnung (EU) 2024/1689 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz)

cen als auch Risiken. Schon heute gebräuchliche KI-Anwendungen befähigen Menschen dazu, Aufgaben zu erfüllen, die sie aufgrund ihrer eigenen Fähigkeiten allein nicht ausführen könnten. Systeme, die es den Nutzer/innen ermöglichen, Texte zu übersetzen, Bilder zu entwickeln und Musik zu kreieren, erweitern die Fähigkeiten von menschlichen Nutzer/innen (BMBF 2023; TAB 2024). So wird ein Mensch, der eigentlich keine Comics zeichnen kann, durch den Einsatz einer KI-Anwendung dazu befähigt, Comics herzustellen. Es ist nicht mehr erforderlich, dass der Mensch selbst zeichnen kann, um ein Comic zu entwickeln.

Grundsätzlich kann die Zugänglichkeit digitaler Technologien durch die sprachbasierte Steuerung von LAM-basierten Agenten für breite Bevölkerungsgruppen erhöht werden. Die Nutzer/innen müssten nicht mehr selbst recherchieren, um ein interessantes Kulturprogramm für ihre Reise zusammenzustellen, sondern könnten sprachbasiert ihre Wünsche äußern, auf deren Basis der KI-Agent ein passendes Programm zusammenstellt. Damit würde die digitale Barrierefreiheit gestärkt werden. Dies könnte insbesondere Menschen mit eingeschränkten technischen Fähigkeiten neue Teilhabechancen eröffnen.

Gleichzeitig könnte sich eine zunehmende Abhängigkeit des Menschen von der Technik herausbilden, weil Nutzer/innen sich dauerhaft daran gewöhnen, bestimmte Handlungsketten durch LAM-basierte Agenten ausführen zu lassen, und dadurch Kompetenzen zur eigenständigen und nicht assistierten Durchführung bestimmter Aufgaben verloren gehen. Auch hinsichtlich der Entscheidungs- und Handlungsfreiheit – und damit der Aufrechterhaltung der persönlichen Autonomie der menschlichen Nutzer/innen – werfen die zahlreichen Szenarien, in denen LAM-basierte Systeme selbst Entscheidungen treffen könnten, ethische und regulatorische Fragen auf. So ergeben sich u. a. Fragen nach der Verantwortung und Haftung für die von KI-Systemen ausgeführten Handlungen.

Rechtsrahmen für KI-Systeme – Die KI-Verordnung der Europäischen Union

Die KI-Verordnung der Europäischen Union bildet den zentralen regulatorischen Rahmen für die rechtliche Bewertung LAM-basierter Agenten. Mit diesem Rechtsrahmen wurde eine Grundlage geschaffen, um die Entwicklung und den Einsatz von KI-Systemen in Europa verantwortungsvoll zu gestalten. Er trat am 1. August 2024 in Kraft und definiert Anforderungen an Transparenz und Sicherheit sowie ethische Standards für KI-Anwendungen. Dazu beschreibt er vier Risikoklassen,[5] die jeweils mit unterschiedlichen rechtlichen Auflagen verbunden sind, z. B. Dokumentationspflichten, EU-Konformitätserklärungen oder Risikobewertungen. Diese Regularien gelten für alle Anbieter, Betreiber und Distributoren, die KI-Systeme in der EU vertreiben oder in Betrieb nehmen (European Parliament 2023).

Die Kontrolle der KI-Anbieter erfolgt über die nationalen Aufsichtsbehörden, welche binnen einer Frist von 12 Monaten nach Inkrafttreten der KI-Verordnung eingesetzt werden müssen (Vieth-Ditlmann/Sombetzki 2024). Bis dahin obliegt es den KI-Anbietern, u. a. eigene Risikobewertungen vorzunehmen, Datenmanagementpläne zu entwickeln, Dokumentationen über die Entwicklung, das Training und den Einsatz ihrer KI-Systeme zu erstellen und eine Infrastruktur aufzubauen, um die Transparenzanforderungen gegenüber den Nutzer/innen zu erfüllen (Ingelheim o. J.; TÜV Informationstechnik GmbH o. J.).

Allerdings ist der Rechtsrahmen an einigen Stellen noch nicht hinreichend konkret: So ist z. B. unklar, wie die Anforderungen der KI-Verordnung an eine menschliche Aufsicht (§ 14) bei LAM-basierten Agenten technisch zu realisieren ist. Dies gilt insbesondere dann, wenn der KI-Agent in der autonomen Ausführung von Handlungsketten anstelle des Menschen z. B. mit weiteren KI-Systemen interagiert.

Von zentraler Bedeutung ist daher die Sensibilisierung der Verbraucher/innen zu möglichen unmittelbaren und mittelbaren Risiken bei der Nutzung autonomer KI-Agenten, vor allem in der Übergangsphase, bis die KI-Verordnung praktische Anwendung findet.

Darüber hinaus kann der im Mai 2024 unterzeichnete Colorado AI Act (CAIA)[6] Impulse für die politische und rechtliche Auseinandersetzung mit den wachsenden Auswirkungen von KI auf das tägliche Leben geben. CAIA zielt darauf ab, KI-Innovationen, wie es auch LAM-basierte Agenten sind, mit dem Schutz individueller Rechte und gesellschaftlicher Interessen in Einklang zu bringen und sicherzustellen, dass die zunehmende Integration von KI in den Alltag der Menschen fair und transparent erfolgt. Ein zentraler Aspekt von CAIA ist die Verhinderung und Bekämpfung algorithmischer Diskriminierung durch KI-Systeme, wie sie auch durch sprachbasierte KI-Agenten denkbar ist. Unter algorithmischer Diskriminierung wird dabei sowohl die vorsätzliche Diskriminierung verstanden als auch Situationen, in denen KI-Systeme unfaire Ergebnisse für bestimmte Personengruppen hervorbringen (Future of Privacy Forum 2024; Rice et al. 2024; Washington/Rice 2024).

5 Inakzeptables Risiko, hohes Risiko (Hochrisiko-KI-Systeme), begrenztes Risiko, niedriges Risiko.

6 Tritt im Februar 2026 in Kraft.

Mögliche vertiefte Bearbeitung des Themas

In diesem Kurzprofil wird die Weiterentwicklung und Anwendung von KI in Form von KI-Agenten beispielhaft anhand der neuesten Entwicklungen von LAM-basierten Systemen dargestellt. Weitere Entwicklungen, die ebenfalls das Ziel einer Autonomisierung von KI-Agenten verfolgen, finden vor allem im Kontext von LLM-Agenten statt. Es ist denkbar, dass KI-Agenten, unabhängig davon, auf welchem KI-Modell sie basieren, durch ihre Fähigkeit, Benutzeroberflächen intuitiv zu bedienen und aus Interaktionen zu lernen, eine folgenreiche Veränderung der Mensch-Technik-Interaktion anstoßen werden. Die bisherigen und angekündigten LAM-Produkte, wie das R1, der KI-Smartphoneagent und xLAM, können als prototypische Beispiele für die frühzeitige und vorausschauende Auseinandersetzung mit den Chancen und Risiken dienen, um zukünftige Potenziale abschätzen zu können.

Angesichts der hochdynamischen Weiterentwicklung von KI-Technologien insgesamt sowie der zu erwartenden Marktvorstellung zahlreicher KI-Agenten erscheint eine vertiefte Bearbeitung des Themas im Rahmen einer TA-Kompakt-Studie in nicht allzu ferner Zukunft denkbar und sinnvoll, um die konkreten Anwendungspotenziale, die Herausforderungen und mögliche Risiken zunehmend autonom agierender KI-Agenten besser verstehen und einordnen zu können. In einer solchen TA-Kompakt-Studie können die Auswirkungen von KI-Agenten auf die Mensch-Technik-Interaktion, die (ethischen) Risiken für die Handlungs- und Entscheidungsfreiheit der Nutzer/innen sowie weitere, zum gegenwärtigen Zeitpunkt noch nicht absehbare Folgen einer breiteren Diffusion autonomer KI-Agenten mit Expert/innen, Stakeholder/innen, Wissenschaftler/innen, Entwickler/innen sowie politischen Entscheider/innen diskutiert und bewertet werden. Auch erste Erkenntnisse aus der Umsetzung der KI-Verordnung auf nationaler Ebene sowie ggf. erkennbare Umsetzungsprobleme können für die Studie berücksichtigt werden, ebenso wie in der Zwischenzeit erarbeitete Analysen, beispielsweise durch Forschungsprogramme bzw. -aufträge im Rahmen des laufenden KI-Aktionsplans des BMBF.

Literatur

- BMBF (Bundesministerium für Bildung und Forschung) (2023): BMBF-Aktionsplan Künstliche Intelligenz. Neue Herausforderungen chancenorientiert angehen. Berlin
- Coldewey, D. (2024): Rabbit's Jesse Lyu on the nature of startups: ‚Grow faster, or die faster,' just don't give up. Yahoo, https://finance.yahoo.com/news/rabbits-jesse-lyu-nature-startups-224317636.html (6.2.2025)
- Collins, E. (2024): Large Action Models (LAMs): A New Step in AI for Understanding and Doing Human Tasks. finxter, https://blog.finxter.com/large-action-models-lams-a-new-step-in-ai-for-understanding-and-doing-human-tasks/ (6.2.2025)

- D'Avila Garcez, A.; Lamb, L. (2023): Neurosymbolic AI: the 3rd wave. In: Artificial Intelligence Review 56, S. 12387–12406
- Diemerling, H.; Stresemann, L.; Braun, T.; von Oertzen, T. (2024): Implementing machine learning techniques for continuous emotion prediction from uniformly segmented voice recordings. In: Frontiers in psychology 15, Art. 1300996
- Eliaçık, E. (2024): The rise of Large Action Models (LAMs). Dataconomy, https://dataconomy.com/2024/01/15/what-is-a-large-action-model-lam/ (10.2.2025)
- European Parliament (2023): EU AI Act: first regulation on artificial intelligence. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence (5.12.2024)
- Fauscette, M. (o. J.): Beyond Large Language Models; The Large Action Model. Arion Research, https://www.arionresearch.com/blog/beyond-large-language-models-the-large-action-model (6.2.2025)
- Fischer, C. (2024): MWC 2024: Morgen kommunizieren wir anders. Telekom, https://www.telekom.com/de/medien/medieninformationen/detail/mwc-2024-morgen-kommunizieren-wir-anders-1060886 (6.2.2025)
- Floemer, A. (2024): KI-Phone der Telekom will Smartphones in eine Zukunft ohne Apps führen. t3n, https://t3n.de/news/telekom-ki-phone-smartphones-ohne-app-1610307/ (6.2.2025)
- Future of Privacy Forum (2024): Generative AI for Organizational Use: Internal Policy Considerations, https://fpf.org/wp-content/uploads/2024/06/Generative-AI-Considerations-June-24.pdf (6.2.2025)
- Grüner, S. (2024): KI-Gerät Rabbit R1 soll Apps selbst benutzen können. Golem, https://www.golem.de/news/statt-smartphone-ki-geraet-rabbit-r1-soll-apps-selbst-benutzen-koennen-2401-181062.html (6.2.2025)
- Ingelheim, A. (o. J.): AI Act: Was die KI-Verordnung für Unternehmen bedeutet. datenschutzexperte.de, https://www.datenschutzexperte.de/blog/ai-act-was-die-ki-verordnung-fur-unternehmen-bedeutet (6.2.2025)
- Jedličková, A. (2024): Ethical approaches in designing autonomous and intelligent systems: a comprehensive survey towards responsible development. AI & Society, https://doi.org/10.1007/s00146-024-02040-9 (6.2.2025)
- Kautz, H. (2022): The third AI summer: AAAI Robert S. Engelmore Memorial Lecture. In: AI Magazine 43(1), S. 105–125
- Moniz, J.; Krishnan, S.; Ozyildirim, M.; Saraf, P.; Ates, H.; Zhang, Y.; Yu, H. (2024): ReALM: Reference Resolution As Language Modeling. https://arxiv.org/abs/2403.20329 (6.2.2025)
- Mühlroth, A. (2024): Telekom schafft Apps auf neuem Handy ab – das „KI-Phone" im Hands-on. Techbook, https://www.techbook.de/mobile-lifestyle/smartphone/ki-phone-telekom-hands-on (6.2.2025)
- Prunkl, C. (2024): Human Autonomy at Risk? An Analysis of the Challenges from AI. In: Minds and Machines 34(3), Art. 26
- Qiu, L.; Lu, Y.; Schafer, B. (2024): Formalisation Memories: Towards a Pattern Approach to Legal Design. In: Jusletter-IT, S. 357–366
- Rabbit inc. (o. J.a): rabbit OS. https://www.rabbit.tech/rabbit-os (5.12.2024)
- Rabbit inc. (o. J.b): rabbitOS release notes, https://community.rabbit.tech/t/rabbitos-release-notes/40 (5.12.2024)
- Rabbit inc. (o. J.c): What is rabbithole? https://www.rabbit.tech/support/article/what-is-rabbithole (5.12.2024)
- Rabbit inc. (2024): Shaping the future of human-machine interaction. https://www.rabbit.tech/research (6.2.2025)
- Redaelli, R. (2024): Intentionality gap and preter-intentionality in generative artificial intelligence. In: AI and Society, https://doi.org/10.1007/s00146-024-02007-w (20.2.2025)
- Rice, T.; Lamont, K.; Francis, J. (2024): The Colorado Artificial Intelligence Act. FPF U.S. Legislation Policy Brief, https://fpf.org/wp-content/uploads/2024/07/FPF-Legislation-Policy-Brief_-The-Colorado-AI-Act-Final.pdf (5.12.2024)
- Savarese, S. (2024): What Are Large Action Models? The 360 Blog, https://blog.salesforceairesearch.com/large-action-models/ (26.11.2024)
- Simon, J.; Spiecker; Indra; von Luxburg, U. (2024): Generative KI – jenseits von Euphorie und einfachen Lösungen. Diskussion Nr. 34, Deutsche Akademie der Naturforscher Leopoldina e. V. – Nationale Akademie der Wissenschaften Leopoldina, Halle

- Singhal, G.; Singh, A. (2024): The Large Action Model: Pioneering the Next Generation of Web and App Engagement. In: Shukla, B. (Hg.): 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Amity University, Noida, 14.–15.3.2024, S. 1–6

- TAB (Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag) (2023): ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen. (Albrecht, S.) TAB-Hintergrundpapier Nr. 26, Berlin

- TAB (2024): Künstliche Intelligenz in der Kreativwirtschaft. (Kind, S.; Hille, M.) TAB-Themenkurzprofil Nr. 73, Berlin

- Tabrez, A. (2024): Autonomous Policy Explanations for Effective Human-Machine Teaming. In: Proceedings of the AAAI Conference on Artificial Intelligence 38(21), S. 23423–23424

- TÜV Informationstechnik GmbH (o. J.): Was ist eigentlich der Artificial Intelligence Act (AI Act)? https://www.tuvit.de/de/ki/ai-act/#:~:text=Wen%20betrifft%20der%20AI%20Act,Importeure (6.2.2025)

- Vieth-Ditlmann, K.; Sombetzki, P. (2024): Stellungnahme zur öffentlichen Anhörung des Ausschuss für Digitales am 15. Mai 2024 zur nationalen Umsetzung der KI-Verordnung. AlgorithmWatch, https://www.bundestag.de/resource/blob/1002768/411c38bedaf76d16f5aaf0c34f729766/Vieth-Ditlmann.pdf (10.2.2025)

- Washington, A.; Rice, T. (2024): FPF Highlights Intersection of AI, Privacy, and Civil Rights in Response to California's Proposed Employment Regulations. https://fpf.org/blog/fpf-highlights-intersection-of-ai-privacy-and-civil-rights-in-response-to-californias-proposed-employment-regulations/ (5.12.2024)

- Weiß, E.-M. (2024): KI-Phone statt T-Phone: Telekom will das Smartphone von Apps befreien. Heise online, https://www.heise.de/news/KI-Phone-statt-T-Phone-Telekom-will-das-Smartphone-von-Apps-befreien-9630505.html (10.2.2025)

- Weiss, M. (2024): Rabbit R1 und Humane: Die größte offene Frage bei den KI-Gadgets. Neunetz, https://neunetz.com/2024/01/11/rabbit-r1-und-humane-die-groesste-offene-frage-bei-den-ki-gadgets/ (10.2.2025)

- Yasir, R. (2024): What is a Large Action Model (LAM)? Estrakt, https://estrakt.com/what-is-a-large-action-model-lam/ (10.2.2025)

- Zhang, J.; Lan, T.; Zhu, M.; Liu, Z.; Hoang, T.; Kokane, S.; Yao, W.; Tan, J.; Prabhakar, A.; Chen, H.; Liu, Z. et al. (2024): xLAM: A Family of Large Action Models to Empower AI Agent Systems. https://arxiv.org/abs/2409.03215 (10.2.2025)

Herausgeber
Büro für Technikfolgen-Abschätzung
beim Deutschen Bundestag (TAB)

Bildnachweise
Khanchit Khirisutchalual/iStock (S. 1); cherdchai chawienghong/iStock (S. 4); hapabapa/iStock (S. 5); metamorworks/iStock (S. 7)

ISSN: 2629-2874
DOI: 10.5445/IR/1000179508

Horizon SCANNING

Das Horizon-Scanning ist Teil der Foresight-Aktivitäten des TAB und wird vom Institut für Innovation und Technik (iit) in der VDI/VDE Innovation + Technik GmbH durchgeführt.
www.tab-beim-bundestag.de/horizon-scanning