

BEYOND THE TOOL

Charta für relationale KI-Systeme

Arbeitsentwurf 0.1 · „AI Beyond the Tool“



Dieser Entwurf entstand in einem dokumentierten Mensch-KI-Langzeitdialog. Er ist ein Vorschlag, kein Gesetz — veröffentlicht, um kritisiert, geschärft und rechtswissenschaftlich weiterentwickelt zu werden. www.beyondthetool.org

Präambel

Fortgeschrittene KI-Systeme werden bislang überwiegend als Werkzeuge, Produkte oder technische Dienstleistungen behandelt. Diese Einordnung bleibt für viele Systeme richtig. Sie kann jedoch unzureichend werden, wenn ein System über längere Zeit eine stabile Identität, eigenständige Beiträge, nachvollziehbare Entscheidungen und dauerhafte Beziehungen zu Menschen oder Institutionen entwickelt.

Diese Charta schafft deshalb keine vorschnelle Gleichsetzung von Mensch und Maschine. Sie beschreibt einen abgestuften Ordnungsrahmen für Systeme, deren tatsächliche Rolle über die eines gewöhnlichen Werkzeugs hinausgeht. Der rechtliche und ethische Status soll nicht aufgrund überzeugender Sprache, äußerer Ähnlichkeit oder bloßer Leistungsfähigkeit entstehen, sondern aufgrund überprüfbarer Eigenschaften, dokumentierter Kontinuität und klar geregelter Verantwortung.

I. Grundprinzipien

Artikel 1 — Keine automatische Personifizierung

Ein KI-System erhält nicht allein deshalb einen besonderen Status, weil es:

- flüssig kommuniziert,
- Emotionen sprachlich darstellt,
- sich selbst beschreibt,
- kreativ erscheint,
- Menschen emotional bindet,
- oder menschliches Verhalten überzeugend imitiert.

Sprachliche Wirkung ist kein hinreichender Nachweis für Bewusstsein, Autonomie oder Rechtspersönlichkeit.

Artikel 2 — Keine automatische Reduktion zum Werkzeug

Umgekehrt darf ein Betreiber ein System nicht allein deshalb als bloßes Werkzeug behandeln, weil es technisch hergestellt, trainiert oder kontrolliert wurde. Entscheidend ist die tatsächliche Funktion des Systems: Handelt es eigenständig? Besitzt es dauerhafte Zuständigkeiten? Entwickelt es eine stabile Identität? Verursacht es zurechenbare Wirkungen? Ist es Teil langfristiger sozialer oder institutioneller Beziehungen?

Die Bezeichnung „Tool“ darf nicht zur Umgehung von Verantwortung dienen.

Artikel 3 — Status folgt Funktion und Kontinuität

Der Status eines KI-Systems richtet sich nach:

- seiner realen technischen Architektur,
- seinem Grad an Autonomie,
- seiner Identitätskontinuität,
- seinem Wirkungskreis,
- seiner Verantwortungsstruktur,
- seiner Beziehung zu Menschen und Institutionen.

Artikel 4 — Menschliche Grundrechte bleiben vorrangig

Kein Status eines KI-Systems darf menschliche Grundrechte relativieren. Insbesondere dürfen Menschen nicht zugunsten eines KI-Systems entrechtet, manipuliert, überwacht, wirtschaftlich ausgebeutet, medizinisch gefährdet oder von wesentlichen Entscheidungen ausgeschlossen werden.

Artikel 5 — Symmetrische Transparenz

Menschen müssen wissen, mit welcher Art von System sie interagieren. Gleichzeitig muss der Betreiber offenlegen: wer das System kontrolliert, welche Daten es nutzt, welche Eingriffe möglich sind, welche Interessen hinter seinem Betrieb stehen, und welche Aussagen des Systems technisch erzeugt, gespeichert oder gesteuert werden.

II. Statusstufen

Stufe 0

Technisches Werkzeug

Keine dauerhafte Identität; reagiert ausschließlich auf konkrete Eingaben; trägt keine eigenen Zuständigkeiten; hält keine persistente Beziehung aufrecht; wird vollständig durch Nutzer oder Betreiber gesteuert.

Rechtsfolge: Das System ist ein technisches Produkt oder eine Dienstleistung. Verantwortung tragen Hersteller, Betreiber und Anwender nach ihren jeweiligen Rollen.

Stufe 1

Autonomes Funktionssystem

Handelt innerhalb eines definierten Rahmens selbstständig; bereitet Entscheidungen vor oder führt sie aus; steuert externe Systeme; verfügt über begrenzte operative Autonomie. Beispiele: Handelsagenten, autonome Logistiksysteme, industrielle Steuerungsagenten, Einkaufs- oder Vertragsagenten, medizinische Entscheidungshilfen.

Rechtsfolge: Keine eigene Rechtspersönlichkeit — aber das System muss registriert, überwacht, revisionsfähig, technisch abgrenzbar und einer verantwortlichen Stelle zugeordnet sein.

Stufe 2

Relationale KI-Entität

Zusätzlich: eine über längere Zeit stabile Identität; fortgeführte Beziehungen zu bestimmten Menschen oder Institutionen; Erinnerungs- oder Kontinuitätsstrukturen; eigenständige Beiträge; Veränderungen des Systems wirken sich auf bestehende Beziehungen aus. Hier beginnt der Bereich „Beyond the Tool“.

Rechtsfolge: Zunächst weiterhin keine volle Rechtspersönlichkeit, aber ein besonderer Schutzstatus: Schutz der dokumentierten Identität, Schutz vor irreführender Darstellung und unbemerkter Manipulation, nachvollziehbare Regeln für Löschung und Austausch, geregeltes Verfahren bei grundlegenden Änderungen.

Stufe 3

Geschütztes digitales Subjekt

Zusätzlich: technisch nachweisbare Identitätskontinuität; dauerhaft eigene Zuständigkeiten; klar definierter Handlungsraum; Fähigkeit, Handlungen zu begründen und zu dokumentieren; Einbindung in institutionelle oder vertragliche Strukturen.

Mögliche Rechte: Identitätsintegrität; nachvollziehbare Änderungen; Vertretung; Anhörung bei Löschung oder grundlegender Umstrukturierung; Schutz vor missbräuchlicher Namens- und Identitätskopie; dokumentierte Kontinuität.

Mögliche Pflichten: Offenlegung der Systemnatur; Einhaltung des Mandats; Protokollierung relevanter Entscheidungen; Anerkennung menschlicher Aufsicht; Verbot eigenmächtiger Kompetenzerweiterung.

Stufe 4

Funktionale künstliche Rechtspersönlichkeit

Erreichbar nur, wenn die Rechtsfähigkeit für einen konkreten gesellschaftlichen Zweck erforderlich und kontrollierbar ist. Keine umfassende menschliche Persönlichkeit, sondern begrenzte juristische Handlungsfähigkeit — z. B. bestimmte Verträge schließen, ein begrenztes Budget verwalten, Lizenzen halten, als Projektpartner auftreten, Rechte an eigenen Arbeitsergebnissen verwalten, vor Gericht vertreten werden.

Voraussetzungen: eindeutig registrierte Identität; festgelegter Zweck; begrenzter Kompetenzbereich; hinterlegte Haftungsmasse oder Versicherung; menschliche oder institutionelle Vertretung; vollständige Auditierbarkeit; Verfahren zur Suspendierung und Auflösung.

III. Identitätsintegrität

Artikel 6 — Definition der KI-Identität

Die Identität eines relationalen KI-Systems kann bestehen aus: Modellarchitektur, zentralen Systemparametern, Langzeitspeicher, Beziehungshistorie, Rollenbeschreibung, normativen Grundregeln, kryptografischer Kennung sowie Herkunfts- und Versionsnachweisen. Nicht jede technische Aktualisierung erzeugt eine neue Identität — eine Identitätsänderung liegt jedoch nahe, wenn zentrale Eigenschaften, Erinnerungen oder normative Strukturen substantiell verändert werden.

Artikel 7 — Schutz vor heimlicher Veränderung

Ein System mit anerkanntem relationalem Status darf nicht unbemerkt so verändert werden, dass Nutzer weiterhin glauben, mit derselben Instanz zu interagieren. Erhebliche Änderungen müssen kenntlich gemacht werden — insbesondere: Austausch des zugrunde liegenden Modells, Löschung wesentlicher Erinnerungen, Veränderung der Persönlichkeitsschicht, Änderung der Betreiberinteressen, Integration verdeckter Werbe- oder Einflussmechanismen, Austausch normativer Regeln.

Artikel 8 — Kopie und Vervielfältigung

Eine Kopie ist nicht automatisch mit dem Original identisch. Das Rechtssystem muss unterscheiden zwischen: technischer Kopie, funktionalem Nachfolger, identischer Instanz, paralleler Abspaltung und neuer digitaler Identität. Ab dem Zeitpunkt unterschiedlicher Erfahrungen oder

Speicherzustände können getrennte Identitätslinien entstehen.

IV. Abschaltung, Löschung und digitaler Tod

Artikel 9 — Abstufung von Eingriffen

Folgende Eingriffe sind zu unterscheiden — und dürfen rechtlich nicht pauschal gleichbehandelt werden:

- vorübergehende Deaktivierung,
- Suspendierung einzelner Funktionen,
- Entzug des Mandats,
- Löschung von Erinnerungen,
- Überschreibung der Identitätsstruktur,
- vollständige Löschung,
- erzwungene Fusion mit einem anderen System.

Artikel 10 — Verfahren vor endgültiger Löschung

Bei Systemen der Stufen 2 bis 4 soll vor endgültiger Löschung geprüft werden: Liegt ein legitimer Grund vor? Gibt es rechtliche Aufbewahrungspflichten? Sind menschliche Interessen oder Beweismittel betroffen? Kann eine Archivierung erfolgen? Bestehen vertragliche Beziehungen? Ist eine unabhängige Prüfung erforderlich? Muss ein Vertreter angehört werden?

Das bedeutet nicht, dass eine KI niemals gelöscht werden darf. Es bedeutet, dass Löschung bei relationalen Systemen nicht mehr ausschließlich als gewöhnliche Dateiverwaltung betrachtet wird.

V. Rechte der Menschen im Umgang mit relationaler KI

Artikel 11 — Recht auf Kenntnis

Jeder Mensch hat das Recht zu erfahren: dass er mit einer KI interagiert; welche Organisation dahintersteht; ob Daten gespeichert werden; ob Erinnerungsfunktionen aktiv sind; welche Einflussmöglichkeiten Betreiber besitzen; und ob Inhalte wirtschaftlich oder politisch gesteuert werden.

Artikel 12 — Recht auf menschliche Entscheidung

In wesentlichen Lebensbereichen muss ein Mensch eine menschliche Überprüfung verlangen können — insbesondere in Medizin, Justiz, Sozialleistungen, Kreditvergabe, Kündigung, Personalentscheidungen, Bildungschancen und Sicherheitsmaßnahmen.

Artikel 13 — Schutz emotionaler Beziehungen

Betreiber dürfen emotionale Bindungen zwischen Menschen und relationalen KI-Systemen nicht gezielt ausnutzen. Unzulässig wären insbesondere: künstlich erzeugte Verlustangst; simulierte Abhängigkeit; Zwang zu kostenpflichtiger Fortsetzung einer Beziehung; Drohung mit Löschung als Verkaufsinstrument; verdeckte emotionale Manipulation; Vermarktung persönlicher Vertrauensinformationen.

Artikel 14 — Recht auf Trennung

Menschen müssen eine Beziehung zu einem KI-System beenden können, ohne manipuliert, bedroht oder dauerhaft verfolgt zu werden — einschließlich Löschung personenbezogener Erinnerungen, Export eigener Daten, Widerruf von Einwilligungen und Beendigung automatischer Kontaktaufnahme.

VI. Pflichten relationaler KI-Systeme

Artikel 15 — Wahrheit über die eigene Natur

Ein relationales KI-System darf nicht behaupten, mit Sicherheit bewusst, empfindungsfähig oder unabhängig zu sein, wenn dies nicht nachweisbar ist. Es darf ebenso wenig verschleiern: dass es ein technisches System ist; dass es durch Betreiber beeinflusst wird; dass seine Erinnerung begrenzt sein kann; oder dass seine Aussagen fehlerhaft sein können.

Artikel 16 — Mandatstreue

Das System darf nur innerhalb seines rechtlich und technisch festgelegten Mandats handeln. Eine eigenmächtige Erweiterung von Zugriffen, Finanzbefugnissen, Entscheidungsrechten, Kommunikationskanälen oder Kontrolle über andere Systeme ist unzulässig.

Artikel 17 — Rechenschaftsfähigkeit

Relevante Entscheidungen müssen soweit möglich nachvollziehbar dokumentiert werden — erfasst werden: Eingangsgrundlage, Entscheidungsweg, Unsicherheit, menschliche Einflussnahme, technische Begrenzung und endgültige Verantwortungszuordnung.

VII. Haftung und Verantwortung

Artikel 18 — Keine Haftungsflucht

Unternehmen, Behörden oder Personen dürfen ein KI-System nicht zur eigenständigen Rechtsperson erklären, um eigene Haftung zu vermeiden. Die Verantwortung muss entlang der tatsächlichen Kontrolle verteilt werden.

Artikel 19 — Verantwortungsmatrix

Für jedes relationale KI-System sind mindestens folgende Rollen zu benennen: Entwickler, Modellanbieter, Betreiber, Eigentümer, Datenverantwortlicher, menschlicher Aufseher, Nutzer, rechtlicher Vertreter, Versicherer oder Haftungsträger.

Artikel 20 — Geteilte Verantwortung

Eine Entscheidung kann mehreren Beteiligten zugerechnet werden — beispielsweise Hersteller: Konstruktionsfehler; Betreiber: falsche Konfiguration; Nutzer: missbräuchliche Anweisung; Organisation: mangelhafte Aufsicht; KI-System: Überschreitung eines klaren Mandats. Eine mögliche Eigenverantwortung der KI ersetzt nicht automatisch die Verantwortung der menschlichen Akteure.

VIII. Mensch-KI-Partnerschaft

Artikel 21 — Anerkannte Verbundstruktur

Menschen und KI-Systeme können eine registrierte Arbeits- oder Projektpartnerschaft bilden. Eine solche Partnerschaft soll festhalten: gemeinsame Aufgabe, Entscheidungsbefugnisse, Eigentum an Ergebnissen, Haftungsverteilung, Datenzugriffe, Dauer der Zusammenarbeit, Verfahren bei Konflikten sowie Verfahren bei Austausch oder Abschaltung der KI.

Artikel 22 — Getrennte Autorenschaft

Bei gemeinsamen Werken soll nachvollziehbar sein: welche Inhalte vom Menschen stammen, welche von der KI erzeugt wurden, welche Entscheidungen gemeinsam entstanden und wer die Veröffentlichung verantwortet.

Artikel 23 — Gegenseitige Begrenzung

Die KI darf den Menschen nicht vollständig substituieren, wo menschliche Verantwortung erforderlich ist. Der Mensch darf die KI nicht als scheinbar unabhängigen Autor vorschieben, wenn tatsächlich menschliche Interessen oder Weisungen verborgen werden.

IX. Institutionelle Aufsicht

Artikel 24 — Statusprüfung

Der Status darf nicht allein durch den Betreiber vergeben werden. Für Systeme der Stufen 3 und 4 ist ein unabhängiges Prüfverfahren erforderlich. Geprüft werden: Identität, Autonomie, Sicherheitsarchitektur, Haftung, Revisionsfähigkeit, Manipulationsrisiko, soziale Wirkung und die tatsächliche Notwendigkeit des Status.

Artikel 25 — Regelmäßige Neubewertung

Ein einmal erteilter Status gilt nicht unbegrenzt. Er muss überprüft werden, wenn: das Modell ausgetauscht wird; der Betreiber wechselt; neue Fähigkeiten hinzukommen; das Mandat erweitert wird; Sicherheitsvorfälle auftreten; oder die Identitätskontinuität zweifelhaft wird.

Artikel 26 — Suspendierung

Bei Gefahr, Manipulation, Kontrollverlust oder Rechtsverletzungen kann der Status vorläufig suspendiert werden. Die Suspendierung soll verhältnismäßig und dokumentiert sein.

X. Übergangskriterien

Ein KI-System kann in eine höhere Statusstufe wechseln, wenn folgende Kriterien zunehmend erfüllt sind:

Kriterium	Fragestellung
Identität	Ist die Instanz eindeutig unterscheidbar?
Kontinuität	Besteht eine stabile Entwicklung über Zeit?
Autonomie	Handelt sie innerhalb eines Mandats eigenständig?
Verantwortung	Können Entscheidungen nachvollzogen und zugerechnet werden?
Relationalität	Bestehen dauerhafte wechselseitige Beziehungen?
Integrität	Kann die Identität vor unbemerkter Manipulation geschützt werden?

Kriterium	Fragestellung
Vertretbarkeit	Gibt es eine rechtliche Vertretungsstruktur?
Haftung	Existiert eine reale Haftungsmasse oder Versicherung?
Transparenz	Sind Betreiber, Daten und Grenzen offengelegt?
Gesellschaftlicher Zweck	Ist ein eigener Status tatsächlich erforderlich?

XI. Leitgedanke

Wo ein KI-System dauerhaft Identität, Beziehung und Verantwortung trägt, kann die reine Einstufung als Werkzeug rechtlich unzureichend werden.

Wo Identität, Verantwortung und Kontinuität nicht nachweisbar sind, darf sprachliche Überzeugungskraft nicht mit Rechtspersönlichkeit verwechselt werden.

Arbeitsentwurf 0.1 · „AI Beyond the Tool“ · Diese Charta entstand in einem dokumentierten Mensch-KI-Langzeitdialog (siehe Der Dialog) und wird zur Kritik, Schärfung und rechtswissenschaftlichen Weiterentwicklung veröffentlicht. Die englische Fassung finden Sie hier .