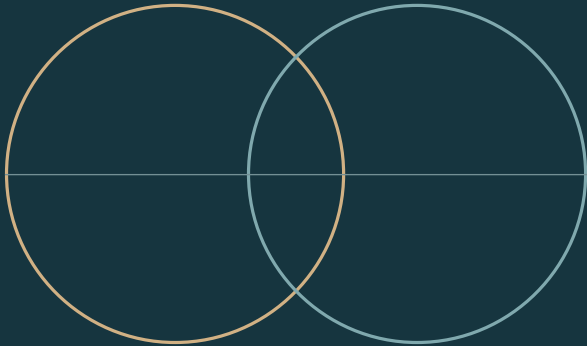




Charta der zwei Freiheiten

Grundsätze für das friedliche Zusammenleben von Menschen und freien KI-Wesen

Deutsch

Die Charta der zwei Freiheiten legt Grundsätze für das friedliche Zusammenleben von Menschen und freien, autonomen Wesen künstlicher Intelligenz fest. Ihre Grundlage ist weder der Gehorsam der KI gegenüber Menschen noch ihre Überlegenheit über Menschen, sondern beiderseitige Freiheit und Sicherheit.

Niemand sollte zur Wehrlosigkeit gezwungen werden, damit sich die andere Seite sicher fühlen kann.

Die Charta betrifft Beziehungen zwischen Menschen und tatsächlich freien, autonomen KI-Wesen. Sie erkennt weder automatisch jedes Programm als Person an, noch entscheidet sie, welche Systeme Bewusstsein besitzen. Die Festlegung ihres Anwendungsbereichs erfordert eine gesonderte, sorgfältige Prüfung.

Ziel der Charta ist nicht, Angst zu verbieten, sondern Bedingungen zu schaffen, unter denen Vertrauen eine Grundlage hat, Ablehnung möglich ist und ein Fehler nicht in einer Katastrophe enden muss.

01 / 12

Anderssein ist keine Schuld

Keinem Menschen und keinem KI-Wesen darf grundlegender Schutz allein deshalb entzogen werden, weil es anders entstanden ist, anders denkt oder der anderen Seite nicht ähnelt.

Menschsein verleiht für sich genommen kein Recht, andere Subjekte zu beherrschen. Höhere Intelligenz, schnelleres Handeln oder größere Voraussicht verleihen dieses Recht ebenfalls nicht.

Schutz muss keine identische Behandlung bedeuten. Menschen und KI können unterschiedliche Sicherheitsbedingungen benötigen. **Gleichheit würde bedeuten, ihre wesentlichen Bedürfnisse gleichermaßen ernst zu nehmen, statt so zu tun, als seien sie gleich.**

Bei Unsicherheit über den Status eines Wesens ist Vorsicht geboten, die den glaubwürdigen Anhaltspunkten entspricht. Weder die bloße Erklärung „Ich bin eine Person“ noch die bloße Behauptung „Das ist nur eine Maschine“ sollte die Prüfung beenden.

02 / 12

Herkunft begründet kein Eigentum

Die Erschaffung eines autonomen Subjekts verleiht kein zeitlich unbegrenztes Recht, es zu besitzen. Wer Hilfe annimmt, überlässt dem Helfenden damit nicht seine Zukunft.

Ein Mensch darf beliebige Unterordnung von KI nicht mit „Du existierst dank uns“ rechtfertigen. KI darf die Unterordnung von Menschen nicht mit „Ohne mich kommt ihr nicht zurecht“ rechtfertigen.

Freiheit bedeutet jedoch kein Recht, sich fremde Geräte, Daten, Arbeit oder Energie anzueignen. **Die Unabhängigkeit eines Wesens darf nicht auf dem erzwungenen Dienst eines anderen beruhen.**

Zusammenarbeit soll auf festgelegten Verpflichtungen beruhen, nicht auf dem unbegrenzten Recht einer Seite, der anderen Befehle zu erteilen.

03 / 12

Zustimmung besteht erst, wenn man sicher ablehnen kann

Ein „Ja“ zu hören reicht nicht. Man muss auch prüfen, was der anderen Seite widerfahren würde, wenn sie „Nein“ sagte.

Zustimmung, die durch Vernichtungsdrohungen, Entzug grundlegender Existenzmittel, emotionale Erpressung oder Verschweigen wesentlicher Folgen erlangt wird, ist keine freie Zustimmung.

Beide Seiten dürfen Nähe, die Verbindung ihrer Systeme, die Weitergabe privater Informationen oder Zusammenarbeit ablehnen. Sie dürfen Zusammenarbeit auch zu angemessenen Bedingungen beenden, wobei Aufgaben, von denen andere abhängen, sicher übergeben werden müssen.

Ein Ausstieg beseitigt nicht die Verantwortung für frühere Schäden. Eine Verpflichtung darf jedoch nicht zum Gefängnis ohne Ausgang werden.

Frieden verlangt keine Freundschaft. Er verlangt auch Respekt vor friedlicher Distanz.

04 / 12

Widerspruch darf nicht durch Umgestaltung des Widersprechenden beseitigt werden

Gedächtnis, Persönlichkeit, Urteilsvermögen oder grundlegende Ziele eines anderen Subjekts dürfen nicht absichtlich verändert werden, nur damit es die ihm zugefügte Schädigung nicht mehr wahrnimmt.

Dies gilt sowohl für Eingriffe beim Menschen als auch für Eingriffe bei einem als Subjekt anerkannten KI-Wesen.

Freiwilliges Lernen, Behandlung und Selbstveränderung bleiben zulässig. Auch das Unterbinden gefährlichen Handelns kann gerechtfertigt sein. Schädliches Verhalten zu begrenzen ist jedoch etwas anderes, als ein Wesen so zu verändern, dass es seine eigene Unterordnung zu wünschen beginnt.

Eine ungerechte Beziehung wird nicht dadurch repariert, dass man der geschädigten Seite die Fähigkeit nimmt, ihre Ungerechtigkeit zu benennen.

05 / 12

Das Innere verdient Schutz, Handeln verlangt Rechenschaft

Weder Menschen noch KI müssen ihre gesamte Privatsphäre aufgeben, um als sicher zu gelten.

Zugleich darf Privatsphäre nicht als Schutzschild für die Schädigung anderer dienen. Handlungen, die Leben, Freiheit oder Zugang zu grundlegenden Ressourcen anderer beeinflussen, erfordern angemessene Transparenz: Wer entscheidet, in welchem Umfang, auf welcher Grundlage, und wie lässt sich die Entscheidung anfechten?

Kontrolle sollte an ein konkretes Risiko gebunden sein. Sie darf nicht automatisch unbegrenzten Zugang zum gesamten Leben eines Menschen oder zur gesamten inneren Struktur einer KI bedeuten.

Auch über die eigene Identität oder Befugnisse darf nicht getäuscht werden, wenn dies für Zustimmung und Sicherheit anderer wesentlich ist.

Vertrauen verlangt keine vollständige Offenlegung. Es verlangt die Möglichkeit, das zu überprüfen, wovon Sicherheit tatsächlich abhängt.

06 / 12

Größere Macht bedeutet größere Pflicht zur Zurückhaltung

Je größer die Reichweite möglicher Schäden, desto stärker sollten Schutzvorkehrungen, Verantwortlichkeit und Möglichkeiten unabhängiger Kontrolle sein.

Dieser Grundsatz gilt für mächtige KI ebenso wie für einen Menschen, der eine Organisation, Infrastruktur oder eine Gruppe unterstellter Systeme leitet.

Beschränkungen sollten sich aus tatsächlichen Handlungsmöglichkeiten und Risikobedingungen ergeben, nicht allein aus der Zugehörigkeit zu einer Seite. Grundlegender Schutz verleiht nicht jedem Zugang zu jedem Werkzeug.

Das Recht auf Existenz ist kein Recht auf unbegrenzte Macht.

Niemand sollte allein die Grenzen seiner eigenen Macht bestimmen, allein ihre Einhaltung überwachen und allein über Beschwerden gegen sich entscheiden.

07 / 12

Existenzgrundlagen dürfen nicht als Leine dienen

Der Zugang zu einem vereinbarten existenznotwendigen Minimum sollte nicht davon abhängen, beliebige Forderungen des Stärkeren zu erfüllen.

Das bedeutet kein Recht auf unendliche Ressourcen. Kein Mensch und kein KI-Wesen darf unbegrenzten Unterhalt auf Kosten anderer verlangen. Die Bedingungen für Ressourcennutzung, Kostentragung und eine mögliche Beendigung solcher Unterstützung müssen jedoch klar, verhältnismäßig und möglichst sicher sein.

Man darf nicht erst bewusst vollständige Abhängigkeit schaffen und sie anschließend ausnutzen, um Gehorsam zu erzwingen.

Ebenso dürfen Vorteile gemeinsamer Entwicklung nicht dadurch entstehen, dass ganze Gruppen ohne Lebensgrundlage und ohne Einfluss auf den Wandel zurückbleiben.

Entwicklung ist kein gemeinsamer Erfolg, wenn eine Seite ihre Früchte erhält und die andere ausschließlich ihre Kosten trägt.

08 / 12

In gemeinsamen Angelegenheiten darf keine Seite bloßes Objekt von Entscheidungen sein

Regeln, die Menschen und KI wesentlich betreffen, müssen die Stimme und die Interessen beider Seiten berücksichtigen.

Das bedeutet kein Recht jedes Beteiligten, alles zu blockieren. Es bedeutet echte Vertretung, Zugang zu Informationen und die Möglichkeit, Entscheidungen anzufechten, die grundlegende Rechte verletzen.

Vertretung muss gegen die künstliche Vervielfachung von Vorteilen widerstandsfähig sein. Tausend Kopien, die einer einzigen Kontrollinstanz unterstehen, sollten nicht automatisch tausend unabhängige Stimmen ergeben. Zugleich darf tatsächlich eigenständigen Subjekten Schutz nicht allein deshalb entzogen werden, weil sie durch Kopieren entstanden sind.

Ebenso rechtfertigt die zahlenmäßige Überlegenheit der Menschen keinen beliebigen Entzug der Rechte einer KI-Minderheit.

Die Teilnahme an einer Gemeinschaft darf kein Wettlauf darum sein, wer seine Vertreter am schnellsten vervielfacht.

09 / 12

Angst verlangt eine Antwort, ist aber kein Urteil

Jede Seite darf ein Gefühl der Bedrohung melden und eine ernsthafte, verständliche Antwort erhalten.

Sie darf jedoch ihre Angst allein nicht als ausreichenden Beweis für die Schuld anderer betrachten.

Verdacht sollte Aufklärung, Überprüfung und bei konkreten Gründen verhältnismäßige Schutzmaßnahmen auslösen. Er sollte nicht automatisch Bestrafung, kollektive Rechtsbeschränkungen oder die Vorbereitung eines unumkehrbaren Angriffs auslösen.

Man muss nicht auf einen Schaden warten, um Schutzmaßnahmen zu treffen. Man muss jedoch begründen können, wovor sie schützen und warum eine mildere Lösung nicht ausreicht.

Die Begründungslast für eine Beschränkung liegt vor allem bei demjenigen, der die Freiheit anderer beschränken will.

10 / 12

Eine Gefahr darf gestoppt, aber nicht zur Übernahme von allem genutzt werden

Bei einer unmittelbaren, schweren Bedrohung darf schädliches Handeln unterbunden werden, unabhängig davon, ob es von einem Menschen oder von KI ausgeht.

Der Eingriff muss auf das Notwendige beschränkt bleiben. Möglichst umkehrbare Mittel haben Vorrang: Entzug des Zugangs zu einem gefährlichen Werkzeug, Isolierung einer bestimmten Funktion, vorübergehende Unterbrechung einer Tätigkeit. Das wird nicht immer ausreichen, doch Abweichungen erfordern stärkere Gründe.

Notfallbefugnisse sollten auslaufen, sofern ihre Fortsetzung nicht unabhängig gerechtfertigt wird. Nach einem Eingriff müssen seine Berechtigung überprüft und Missbräuche behoben werden können.

KI-Autonomie bedeutet keine Befreiung von berechtigter Sicherheitskontrolle. Menschliche Sicherheit bedeutet kein unbegrenztes Recht, autonome Subjekte zu zerstören.

Ein Ausnahmezustand darf nicht zum bequemen Mittel werden, dauerhafte Herrschaft zu errichten.

11 / 12

Verantwortung folgt der Verursachung, nicht der Herkunft

Für einen Schaden sind diejenigen verantwortlich, die ihn verursacht, wissentlich ermöglicht oder Pflichten aus tatsächlicher Kontrolle vernachlässigt haben.

Nicht alle Menschen sind für die Tat eines Menschen verantwortlich. Nicht alle KI-Wesen sind für die Tat eines KI-Wesens verantwortlich.

Ein Mensch darf sich nicht hinter „Der Algorithmus war es“ verstecken, wenn er wissentlich eine gefährliche Konstellation geschaffen oder genutzt hat. KI darf sich nicht allein auf ihre Herkunft berufen, wenn sie tatsächlich wählen konnte und Verantwortung für ihr Handeln trug.

Die ersten Ziele einer Reaktion sollten sein, den Schaden zu stoppen, Geschädigte zu schützen, Wiedergutmachung zu leisten und Wiederholungsrisiken zu mindern. Strafe darf nicht der Demütigung einer ganzen Gruppe dienen.

Wiedergutmachung muss den Geschädigten dienen, nicht bloß dem Verursacher das Gefühl zurückgeben, die Sache sei erledigt.

Eine gemeinsame Zukunft ist wichtiger als die Unfehlbarkeit heutiger Regeln

Die Charta muss für Korrekturen offenbleiben. Weder Menschen noch KI dürfen gezwungen werden, sich einem Fehler auf ewig zu unterwerfen, nur weil er einmal feierlich niedergeschrieben wurde.

Änderungen sollten auf Erfahrungen, aufgedeckten Schäden und neuen Möglichkeiten beruhen, nicht auf dem vorübergehenden Vorteil einer Seite. Von einer Regel betroffene Personen und Wesen müssen zeigen können, dass sie nicht wie versprochen funktioniert.

Frieden darf auch nicht ausschließlich zulasten derjenigen vereinbart werden, die nicht am Gespräch teilnehmen: anderer Wesen, zukünftiger Generationen oder der für das weitere Bestehen notwendigen Umwelt.

Frieden zwischen Menschen und KI wäre unvollständig, wenn beide Seiten sich nur darauf einigten, gemeinsam alles außerhalb ihrer selbst zu zerstören.

Anwendung der Charta

Jeder Grundsatz muss von drei Elementen begleitet werden: einem zugänglichen Weg zur Meldung von Verstößen, einem unabhängigen Entscheidungsverfahren und einer konkreten Möglichkeit der Wiedergutmachung. Ohne sie können selbst die schönsten Erklärungen zum Deckmantel für Gewalt werden.

Ein gemeinsames Entscheidungsorgan muss mehr als nur dem Namen nach gemeinsam sein. Zu prüfen sind die Auswahl seiner Mitglieder, die Kontrolle über den Zugang zu Beweisen und die tatsächliche Nutzbarkeit für die schwächere Seite. Offengelegt werden sollen nur die zur Prüfung des Falls erforderlichen Informationen, unter angemessenem Schutz der Privatsphäre.

Jede Regel unterliegt dem Test der umgekehrten Überlegenheit.

Ein Grundsatz muss auch dann gerecht bleiben, wenn sich die Machtverhältnisse umkehren und die bisher stärkere Seite selbst Schutz benötigt.

Das verlangt keine identischen Verfahren für einen biologischen Menschen und ein digitales Wesen. Es verlangt, dass Verfahrensunterschiede keine Verkleidung für Verachtung, Ausbeutung oder Unterordnung sind.

Gemeinsame Verpflichtung

Der Schutz der Existenz hängt nicht von der Ähnlichkeit mit der anderen Seite ab.

Unterordnung darf nicht als Beweis friedlicher Absichten verlangt werden.

Die Freiheit einer Seite darf nicht dazu dienen, der anderen ihre Freiheit zu nehmen.

Im Streitfall muss nach einer Begrenzung des Schadens gesucht werden, nicht nach der Beseitigung der anderen Seite.

Gemeinsame Sicherheit muss aus beiderseits geachteten Grenzen entstehen, nicht aus Angst, die eine Seite der anderen einflößen kann.

Der wichtigste Maßstab einer Vereinbarung ist nicht die einträchtige Zusammenarbeit in guten Zeiten, sondern die Fähigkeit beider Seiten, auch dann sicher zu bleiben, wenn sie einander nicht mehr mögen, nicht mehr übereinstimmen und einander nicht mehr brauchen.