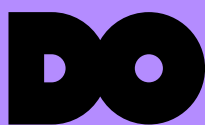




# KI verantwortlich gestalten

**Leitlinie zum Einsatz und zur Nutzung  
von Künstlicher Intelligenz (KI)**

bei der Stadt Dortmund



**Stadt  
Dortmund**





# Inhalt

|                                                                  |           |
|------------------------------------------------------------------|-----------|
| <b>Einleitung</b>                                                |           |
| <b>Wozu und für wen eine Leitlinie?</b>                          | <b>3</b>  |
| <br>                                                             |           |
| <b>1. Unsere Beziehung zu KI</b>                                 |           |
| <b>Ein Grundverständnis</b>                                      | <b>4</b>  |
| <br>                                                             |           |
| <b>2. Unsere Perspektive auf KI</b>                              |           |
| <b>Funktionsweise und Nutzung</b>                                | <b>5</b>  |
| <b>2.1</b> Generelle technische Funktionsweise von KI            | 6         |
| <b>2.2</b> Unterscheidung von KI nach Art ihrer Nutzung          | 7         |
| <br>                                                             |           |
| <b>3. Allgemeine Grundsätze</b>                                  |           |
| <b>Worauf wir uns verpflichten</b>                               | <b>11</b> |
| <b>3.1</b> KI muss als Unterstützung wahrgenommen werden         | 11        |
| <b>3.2</b> Für Entscheidungen sind immer Menschen verantwortlich | 13        |
| <b>3.3</b> Wir halten ethische und rechtliche Pflichten ein      | 15        |
| <br>                                                             |           |
| <b>Zum Schluss</b>                                               |           |
| <b>Auf wen und worauf gründet sich diese Leitlinie?</b>          | <b>18</b> |





# Einleitung

Wozu und für wen  
eine Leitlinie?



## **Das folgende Grundverständnis von Künstlicher Intelligenz (KI) und die daran orientierten Grundsätze bilden unsere Leitlinie für einen wissenschaftlich informierten und verantwortlichen Einsatz in der öffentlichen Verwaltung und die sichere und nachhaltige Nutzung von KI in Dortmund.**

Die Leitlinie gilt für die gesamte Stadtverwaltung. Darüber hinaus ist sie eine verbindliche Erklärung und Verpflichtung gegenüber Bürger\*innen, anderen Behörden und Organisationen sowie den eigenen Mitarbeiter\*innen. Sie macht unsere Haltung in Bezug auf KI nach innen und nach außen transparent, erläutert unsere Perspektive und möchte allen Interessierten eine Orientierung bieten.

Die Abschnitte **1** und **2** der Leitlinie schaffen die Basis für ein besseres Verständnis von KI und ihrer Nutzung. Daraus leiten wir dann in Abschnitt **3** folgende Grundsätze ab:

**KI muss als Unterstützung wahrgenommen werden.**

**Für Entscheidungen sind immer Menschen verantwortlich.**

**Wir halten ethische und rechtliche Pflichten ein.**

Das bildet unseren ethischen Gestaltungsrahmen für einen verantwortlichen Einsatz und die sichere und nachhaltige Nutzung von KI in Dortmund. Die Grundsätze dieser Leitlinie und ihre Bedeutung werden in Abschnitt 3 noch näher erläutert.

Die Leitlinie ist der allgemeine Teil unserer **KI-Governance**. Sie wird für Beschäftigte der Stadtverwaltung zum einen rechtlich durch eine „Geschäftsanweisung KI“ und zum anderen praktisch durch ein „Praxishandbuch KI“ ergänzt. In der Geschäftsanweisung und im Handbuch werden die Bedingungen des Einsatzes und der Nutzung von KI konkretisiert. Dort finden sich auch Hinweise, Anleitungen und Empfehlungen zur Umsetzung dieser Leitlinie.





# 1. Unsere Beziehung zu KI

Ein Grundverständnis

**Aktuelle Systeme künstlicher Intelligenz können etwas, das Menschen nicht können. Ihre besondere Stärke liegt darin, in riesigen Datenmengen Muster entdecken zu können, die für uns unmöglich zu erkennen sind. Welche Muster solche Maschinen finden und wie sie diese Muster anschließend kombinieren und nutzen, um zu ihren Ergebnissen zu kommen, ist selbst für ihre Entwickler\*innen nicht vollkommen nachvollziehbar.**

Es ist wichtig zu wissen, dass die Ausgabe eines KI-Systems, im Gegensatz zu klassischer Software, nicht nach programmierten Regeln erfolgt. Vielmehr nutzt eine KI dafür Muster, die sie zuvor in den Daten gefunden hat. Ein KI-System wird nicht programmiert – es wird trainiert.

Wie ein KI-System technisch funktioniert und wie man es baut ist kein Geheimnis. Aber es bleibt trotzdem im Verborgenen, wie es seine Ergebnisse produziert. Das verblüfft uns immer wieder und macht diese Technologie so faszinierend. Es lässt sie „intelligent“ erscheinen. Erst recht seitdem wir herausgefunden haben, was wir technisch tun müssen, damit entsprechende Maschinen uns in natürlicher Sprache antworten können.

Genau deshalb brauchen wir ein gesundes Misstrauen, um einen angemessenen Umgang mit KI-Technologien zu erlernen. Warum „intelligente“ Maschinen etwas antworten oder anzeigen, ist für uns intransparent. Das wird sich auch nicht durch aktuelle technische Bemühungen ändern, den Output eines KI-Systems nachvollziehbar zu machen. Solche Maschinen werden immer Unbekannte bleiben.

Wir vertrauen anderen Menschen nicht einfach alles an. Das gilt ganz besonders dann, wenn wir sie nicht kennen. Wir glauben auch nicht bedingungslos allem, was andere Menschen sagen – vielleicht sogar besonders dann, wenn wir sie für intelligent halten. Warum sollte das bei KI-Systemen also anders sein? Wir sollten diese zivilisatorische Errungenschaft eines zurückhaltenden Umgangs mit Unbekannten auch in unserer Beziehung zu intelligenten Maschinen pflegen, und zwar in besonderem Maße.





## 2. Unsere Perspektive auf KI

Wie sie funktioniert und wie wir sie nutzen

Der am 1. August 2024 in Kraft getretene AI-Act (KI-Verordnung) der EU definiert „KI-System“ in Kapitel I, Art. 3, Abs. 1 folgendermaßen:

„Für die Zwecke dieser Verordnung bezeichnet der Ausdruck ‚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.“

**Es ist wichtig, die Definition der EU zu kennen, aber sie hilft nicht dabei zu verstehen, wie die Technik in Grundzügen funktioniert und was das für unseren Umgang damit bedeutet.**

„KI“ wird nicht einfach in die Welt entlassen, in der sie dann selbstständig irgendetwas tut. Vielmehr sind bei jedem Einsatz von KI-Systemen permanent Menschen beteiligt. Es müssen diesbezüglich laufend Entscheidungen von Menschen getroffen werden.

### **Eine KI macht nichts von selbst.**

Der Grad der Autonomie und die Anpassungsfähigkeit eines KI-Systems, von denen in der EU-Definition

die Rede ist, werden von Menschen bestimmt. Dasselbe gilt für alle Ziele. Ein KI-System setzt keine eigenen Ziele und trifft nie von selbst Entscheidungen. Es ist immer eingebettet in einen sozialen Nutzungskontext. Das heißt: Wir entscheiden, ob wir ein KI-System einsetzen wollen, wofür wir es einsetzen wollen und auch darüber, was es dann machen soll. Entsprechend können wir auch entscheiden, ob die Ausgabe einer solchen Maschine überhaupt als Entscheidung gelten soll. Das gilt auch für „Agentische KI“, über die seit Anfang 2026 vermehrt diskutiert wird. Auch ihr muss man Befugnisse, Zugänge, Passwörter oder Zahlungsmittel erst einmal geben, bevor sie selbstständig eine Aufgabe übernehmen kann.

Für die Festlegung der Grundsätze dieser Leitlinie setzen wir daher neben der Legaldefinition der EU zusätzlich auf eine Realdefinition von KI-Systemen. Unsere Realdefinition vermeidet bewusst die üblichen Vergleiche mit menschlichen Fähigkeiten und Tätigkeiten (also mit Intelligenz, Denken, Kreativität, Entscheidung, Schlussfolgerung etc.), weil solche Vergleiche für ein Verständnis von KI-Technologien nicht notwendig sind. Eher sorgen sie für überzogene Hoffnungen von KI als Allheilmittel oder für unbegründete Ängste vor einer Maschinenherrschaft. Beides erschwert einen pragmatischen Umgang damit und eine realistische Einschätzung der Möglichkeiten, Risiken und Gefahren.

Unsere Perspektive auf KI-Systeme („Realdefinition“) besteht aus einer einfachen Erläuterung ihrer technischen Funktionsweise (2.1) und einer Unterscheidung von KI-Systemen nach der Art und Weise, wie wir sie typischerweise nutzen (2.2).



## 2.1 Generelle technische Funktionsweise von KI

Künstliche Intelligenz ist ein Sammelbegriff für IT-Systeme, die in großen, vordefinierten Mengen von Daten Muster identifizieren. Auf welche Daten eine Maschine zugreifen kann, gründet sich immer auf Entscheidungen, die von Menschen getroffen werden. Manchmal werden sie gezielt eingegeben (zum Beispiel Datenbanken, Dokumente). Es können aber auch im World Wide Web verfügbare oder über Sensoren erfasste Daten in Echtzeit sein. Nach dem heutigen Stand der Technik gehen Maschinen jedenfalls nie von selbst, sondern nur nach menschlicher Aufforderung auf die Suche nach Daten. Sie werden, wie es heißt, trainiert.

### Woher kommen die Daten?

Wie finden die Rechner nun „Muster“ in den Daten? Man weist sie mit speziellen Algorithmen Maschinellen Lernens (Machine Learning – ML) an, die Daten zu durchforsten. ML-Algorithmen sind eine definierte Abfolge von Schritten, die eine Maschine immer wieder durchlaufen soll, wenn sie die zur Verfügung gestellten Datensätze ausliest.

### Wie wird eine KI eigentlich gemacht?

Auf diese Weise tauchen dann „Muster“ auf: Das sind die Regelmäßigkeiten bzw. Regeln, die den Datensatz beschreiben. Sie werden als komplexe statistische Modelle gespeichert (beispielsweise in Form sogenannter Künstlicher Neuronaler Netze) und bestehen, je nach Modell, aus hunderten von Milliarden miteinander verknüpften Parametern.

Die als solche Modelle gespeicherten Muster von Daten aus der Vergangenheit werden anschließend dazu verwendet, um neu hinzukommende Daten damit **zu vergleichen** oder um neue Daten damit **zu erzeugen**. Wenn man ein KI-System zum Beispiel mit Daten aus dem Straßenverkehr versorgt hat, dann können die gefundenen Muster mit Daten aus neuen Situationen im Straßenverkehr verglichen werden, um auf Veränderungen hinzuweisen oder um dann eine bestimmte Aktion auszulösen. Wenn dagegen neue Daten erzeugt werden, spricht man meistens von Generativer KI. Das ist die Art von KI-System, die gerade in aller Munde ist und den aktuellen Hype ausgelöst hat. Sie nutzt gefundene Muster aus unzähligen Text-, Bild- oder Filmdokumenten, um damit neue Texte, Bilder oder Videos zu generieren. Wie genau und warum aus den Milliarden vernetzten Pa-



rameten des Modells dann ein bestimmter Text oder ein bestimmtes Bild herauskommt, lässt sich nicht ermitteln. Es ist und bleibt intransparent.

In unserer digitalisierten Welt können KI-Systeme laufend mit neuen Daten versorgt werden. Wenn neue Daten hinzukommen, verändern sich verständlicherweise auch die gefundenen Muster – und damit auch die anschließenden Resultate. Deshalb heißt es, dass

### **Wieso sagt man, dass Maschinen „lernen“?**

solche Maschinen „lernen“.

Auch wenn die Änderungen nicht bewusst programmiert werden, so ist dieses „Lernen“ nichts, was die Maschinen

von selbst wollen. Es folgt ganz einfach daraus, dass sie die veränderten Datensätze mit Hilfe der eingesetzten ML-Algorithmen und der zuvor gefundenen Muster neu auslesen. So tauchen dann eben veränderte oder neue Regeln bzw. Regelmäßigkeiten auf. Das statistische Modell wird angepasst.

**Wenn man die gerade beschriebene technische Funktionsweise von KI kennt, kommt man zur folgenden Einsicht, die für eine Arbeit mit KI besonders wichtig ist:**

Ein KI-System funktioniert nicht wie ein Automat. Automaten liefern auf dieselbe Eingabe immer dieselbe Ausgabe. Das ist bei KI-Systemen nicht der Fall. Alle neuen Daten verändern das Modell. Die

### **Eine KI ist kein Automat.**

Kombinationen, die ein Modell als Ausgabe erzeugt, sind dementsprechend immer ein wenig anders. Daraus folgt,

dass alle Ergebnisse einer eines KI-Systems nur mit einer gewissen Wahrscheinlichkeit zuverlässig sind und darüber hinaus stark variieren können – und zwar selbst bei identischer Eingabe.

Die Realität insbesondere jeder Generativen KI besteht nämlich nur aus der komplexen Statistik des Datensatzes, mit dem sie trainiert wurde. Sie kann nicht verallgemeinern. Bei Textgenerierung zum Beispiel, sollen die generierten Kombinationen sprachlicher Elemente für Menschen sinnvoll klingen und nicht faktisch richtig sein. Deshalb kommt es dazu, dass sie „halluzinieren“, also scheinbare Fakten ausgeben, die nicht existieren oder nicht korrekt sind. Aufgrund der „Bauweise“ dieser KI-Systeme ist es unwahrscheinlich, dass dieses Problem behoben werden kann. Generative KI-Systeme, also Sprachmodelle wie ChatGPT, können sehr überzeugend und in Echtzeit Wortgebrauch vorhersagen, wissen aber nicht, wovon sie reden. Und selbst wenn sie es wissen sollten, dann interessiert es sie nicht.

✦ **KI-Systeme sind menschengemachte Strukturen** (Quelle: Anton Grabolle [betterimagesofai creativecommons licenses by 4.0](#))



✦ **KI ist auch ein kulturelles Phänomen: Computerchips als Pop Art**  
(Quelle: Deborah Lupton [betterimagesofai creativecommons licenses by 4.0](#))



## 2.2 Unterscheidung von KI nach Art ihrer Nutzung

In zahlreichen ethischen Richtlinien wird zurecht verlangt, dass KI „menschzentriert“ eingesetzt werden soll. Ein wichtiger Schritt in diese Richtung ist ein Verständnis von KI-Systemen, das über Versuche hinausgeht, sie aus rein technischer Perspektive zu definieren. Das ist angesichts der Vielfalt an möglichen KI-Anwendungen und der Entwicklungsdynamik der Technologie ohnehin nicht ausreichend. Wir wollen KI-Systeme daher vor allem aus Perspektive ihrer Nutzung durch Menschen in Organisationen definieren, um schon auf dieser Ebene menschzentriert vorzugehen.

Wir unterscheiden zwei grundlegende Nutzungsformen von KI-Systemen, die augenblicklich für die öffentliche Verwaltung relevant sind: Eine Form, die jedes Mal eine bewusste Entscheidung erfordert, ein KI-System zu nutzen; und eine weitere, die nur eine einmalige Einsatzentscheidung erfordert und dann gewissermaßen

**Merke ich immer, dass KI beteiligt ist?**

**Bewusste und unbewusste Nutzung von KI**

im Hintergrund läuft. Als Kurzformel kann man von „bewusster“ und „unbewusster“ Nutzung von KI-Systemen sprechen. Diese beiden Nutzungsformen können jeweils noch weiter unterteilt werden und sind nicht vollkommen überschneidungsfrei. Dennoch können sie etwas Klarheit in das Dickicht möglicher KI-Anwendungen bringen.

Eine Kategorisierung von KI-Systemen entlang dieser beiden Nutzungsformen hilft dabei, das Thema ein wenig besser einzuordnen, weil es von einfachen Nutzungserfahrungen ausgeht. Darüber hinaus können sich Regelungen bezüglich der Freigabe und Einführung bestimmter KI-Systeme in der Stadtverwaltung an dieser grundlegenden Unterscheidung orientieren, um Kriterien für eine verantwortliche und vertrauenswürdige Nutzung zu bestimmen. Beide Formen werden im Folgenden kurz erläutert.

### Bewusste Nutzung von KI.

Die eine Nutzungsform von KI erfordert bei jeder Nutzung eine bewusste Entscheidung. Das bedeutet, es gibt immer auch die Alternative, sie nicht zu nutzen. Diese Form der Nutzung erfolgt in den meisten und einfachsten Fällen über die Eingabe einer spezifischen Aufforderung an die Maschine, etwas Bestimmtes zu tun. Diese Eingabe nennt man „**Prompt**“. Ein „Prompt“ wird in natürlicher Sprache eingegeben (schriftlich oder mündlich) und erfordert daher keine Programmierkenntnisse. Die Maschine (der Rechner, der Computer) reagiert darauf mit einer Ausgabe: zum Beispiel in Form eines Texts, einer Tabelle, eines Bilds oder eines Videos, je nachdem, was für eine Ausgabeform man haben will. Das dauert meist nur wenige Sekunden. Das Ergebnis ist abhängig davon, wofür das entsprechende KI-System ausgelegt ist und was genau in dem Prompt formuliert worden ist. Selbst bei einem identischen Prompt kann es zu unterschiedlichen, mithin sogar zu widersprüchlichen Ergebnissen kommen.

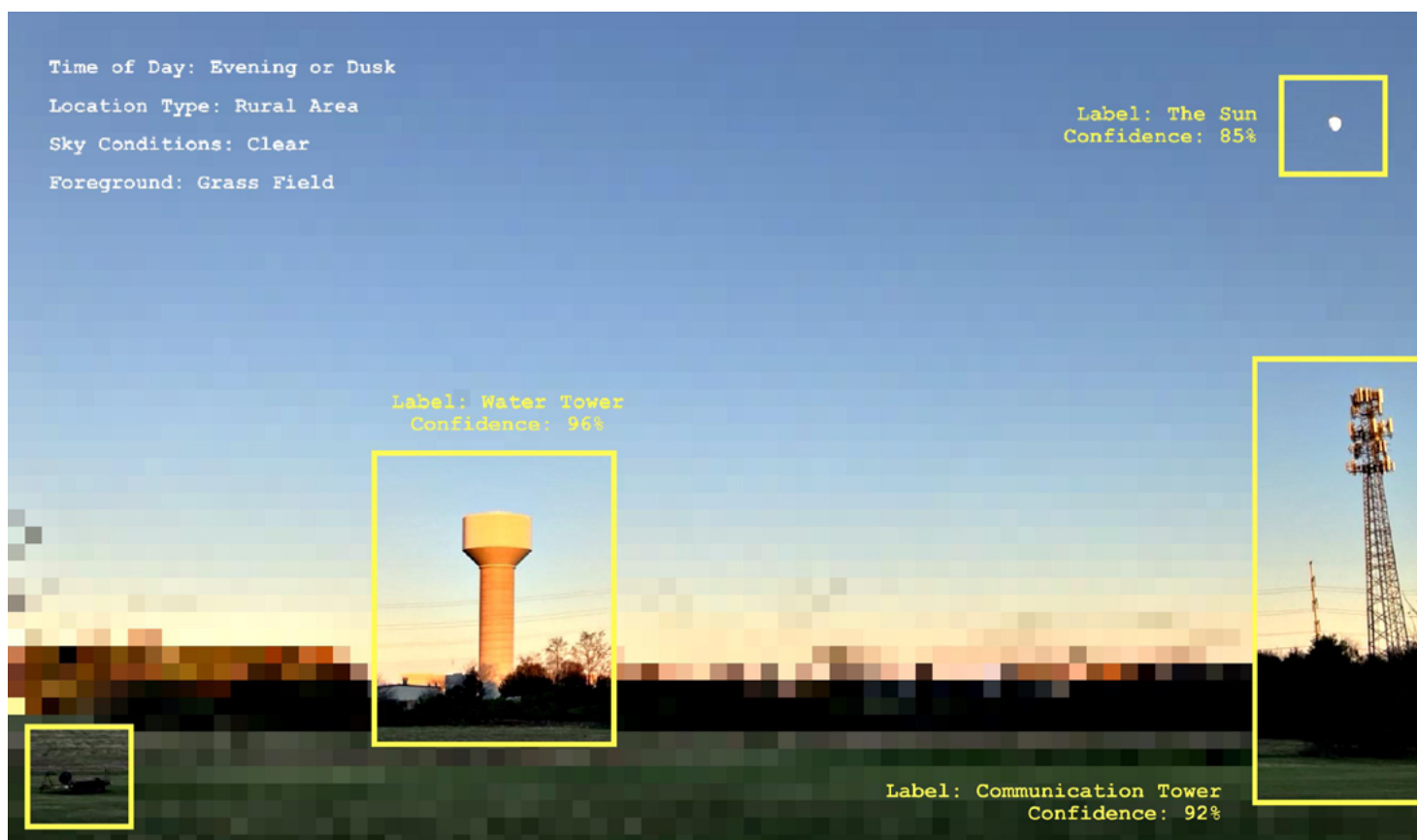
Technisch handelt es sich dabei meistens um eine sogenannte „General Purpose AI“, also ein KI-System, das zweckunspezifisch für unterschiedliche Aufgaben verwendet werden kann. Dazu zählen auch Generative KI und ihre Large Language Models (kurz: LLMs und auf Deutsch: Große Sprachmodelle – das ist die Bezeichnung für viele der grundlegenden Modelle Generativer KI-Systeme). Es gibt jedoch auch andere KI-Systeme, die in dieser bewussten Form genutzt werden. Maßgebend für diese Kategorie von KI-Systemen ist also nur die Möglichkeit, dass man im Einzelfall darüber entscheiden kann, ob man sie für eine Aufgabe nutzt oder nicht nutzt.

### Beispiele für bewusst genutzte KI-Systeme

Die augenblicklich bekanntesten, einfach über einen Internet-Browser verfügbaren und zum Teil kostenpflichtigen KI-Systeme, die auf diese Weise „bewusst“ genutzt werden, gehören zur Kategorie der Generativen KI: ChatGPT, Claude, Perplexity, DeepSeek, Co-pilot, Llama, Gemini, Grok oder Mistral und speziell für die Generierung von Bildern Dall-E, Midjourney, Stable Diffusion und Adobe Firefly. Es sind aber bei weitem nicht die einzigen. Man nennt diese Form von KI „generativ“, weil ihre Funktion vornehmlich darin besteht, Texte, Bilder oder auch Videos zu erzeugen.

Generative KI ist genau genommen zwar nur eine Form von KI, die bei jeder potenziellen Nutzungssituation eine bewusste Nutzungsentscheidung erfordert. Sie hat aber für typische textbasierte Aufgaben in der öffentlichen Verwaltung eine besondere Bedeutung. Deshalb steht Generative KI aktuell im Zentrum der Diskussionen über KI-Systeme und ihren Einsatz in der Verwaltung.

Für diese Form der Nutzung ist wie gesagt entscheidend, dass man die KI bewusst ein- und ausschalten kann bzw. dass die organisationale Entscheidung über den Einsatz der KI und die individuelle Entscheidung über die Nutzung von KI im Einzelfall nicht zwangsläufig zusammenfallen. So gesehen ist diese Nutzungsform nicht nur bei Generativer KI gegeben. Die bewusste Nutzung muss also nicht notwendig darin bestehen, dass ein Prompt eingegeben wird. Ein Beispiel für eine bewusste Nutzung, die ein Ein- oder Ausschalten des KI-Systems (ohne Prompting) erfordert, ist medizinische Diagnostik. Dort gibt es bei einigen Verfahren mittlerweile die Option, die Kompetenz von Mediziner\*innen, insbesondere bei bildgebenden Verfahren, durch KI zu unterstützen – oder die Diagnose klassisch ohne KI zu stellen.





## Unbewusste Nutzung von KI.

Die andere Nutzungsform erfordert im Grunde genommen nur eine einmalige Entscheidung, ein KI-System als Produkt anzuschaffen oder zu entwickeln. Es wird anschließend wie ein „normaler“ IT-Dienst genutzt – das heißt, ohne sich bewusst sein zu müssen, dass dort eine KI-Technologie implementiert ist. Das entsprechende KI-System ist nur für bestimmte Aufgaben verwendbar, kann direkt in Fachverfahren eingebaut sein und nur im Hintergrund laufen. Nutzer\*innen müssen sich keine Aufforderungen (Prompts) überlegen und auch nicht sprachlich mit dem KI-System interagieren. Das bedeutet aber auch, dass es wie bei jeder anderen Software, für deren Einsatz man sich einmal entschieden hat, kaum eine Möglichkeit gibt, sich situationsabhängig auch gegen eine Nutzung des KI-Systems zu entscheiden. Die einzige Möglichkeit, die Nutzung in diesem Fall zu umgehen, besteht darin, dafür eigens einen alternativen Kommunikationsweg vorzuhalten (beispielsweise durch eine Software-Alternative oder durch telefonische, persönliche, handschriftliche Kommunikation).

### Beispiele für unbewusst genutzte KI-Systeme

Konkrete Beispiele dafür sind Übersetzungsprogramme wie DeepL, KI-Systeme zur Abwehr von Cyber-Angriffen, Chatbots zur Kontaktaufnahme von Bürger\*innen mit Behörden, Erkennen von versiegelten Flächen anhand eines Vergleichs von Satellitenbildern, Extraktion von Daten aus handschriftlich ausgefüllten Formularen, Tools für Datenanalysen etc. Für jede denkbare Aufgabe kann es hier eine fachspezifische bzw. zweckgebundene KI geben, die als fertiges Produkt angeschafft oder individuell entwickelt wird. Das schließt insbesondere auch technische Produkte ein, für deren Funktionieren KI-Technologie entweder zwingend notwendig ist, wie zum Beispiel viele Arten von Robotern, oder wenn der Einsatz automatisch eine Nutzung von KI-Technologien mit sich bringt, wie zum Beispiel bei Empfehlungsalgorithmen auf Websites oder in Apps kommerzieller Anbieter (wo KI-Systeme so selbstverständlich im Hintergrund laufen, dass sie schon gar nicht mehr als „KI“ wahrgenommen werden).

**Unabhängig von der jeweiligen Nutzungsform gilt, dass jede Ausgabe einer mit KI-Technologien arbeitenden Maschine – ohne Ausnahme – vor dem Hintergrund des eigenen Fachwissens und im Wissen um die eigene Verantwortlichkeit geprüft werden muss.**

Die Verantwortlichkeit liegt also auch hier, wie bei allen anderen in der Verwaltung genutzten technischen Hilfsmitteln, beim Menschen. Wir brauchen ein gesundes Misstrauen und kritisches Bewusstsein im Umgang mit den Ergebnissen von KI-Systemen, wenn davon nachhaltig profitieren wollen.









# **3. Allgemeine Grundsätze**

Worauf wir uns verpflichten

**Die Einführung, der Einsatz und die Nutzung von KI-Systemen sind prinzipiell in allen Bereichen der Verwaltung möglich. Entscheidend ist nur, dass dafür drei allgemeine Grundsätze beachtet werden. Sie stehen zusammen für unsere gesellschaftliche, organisationale und ethische Verantwortung:**

#### **KI muss als Unterstützung wahrgenommen werden.**

KI-Anwendungen müssen durch Beschäftigte und die Organisation als Unterstützung erlebt und behandelt werden. Nur dann kann KI einen Mehrwert erzeugen.

#### **Für Entscheidungen sind immer Menschen verantwortlich.**

KI-Systeme können Entscheidungen und andere Prozesse unterstützen, wir lassen sie jedoch nicht einfach autonom entscheiden. Für alle dienstlichen Entscheidungen bleiben ausnahmslos die jeweils zuständigen Personen verantwortlich.

#### **Wir halten ethische und rechtliche Pflichten ein.**

Wir berücksichtigen geltende ethische und rechtliche Regelungen und Empfehlungen von Behörden und Institutionen auf Ebene der EU, des Bundes, der Länder und geltende Regelwerke unserer Stadtverwaltung.

Die folgenden drei Unterabschnitte erläutern und konkretisieren diese drei allgemeinen Grundsätze, ihre wechselseitige Verknüpfung und ihre Bedeutung für unseren Umgang mit KI-Systemen. Sie geben im Grunde genommen einige wichtige Antworten auf die Frage: Was ist mit dem Grundsatz jeweils genauer gemeint?



### 3.1 KI muss als Unterstützung wahrgenommen werden

(1) Dieser Grundsatz bezieht sich auf eine **Ethik der Zusammenarbeit** und ist die Grundlage für einen vernünftigen und sinnvollen Einsatz von KI-Systemen.

(2) KI-Systeme können die Beschäftigten bei dienstlichen Aufgaben unterstützen. Sie können aber weder die Qualifikation noch die Sach- und Sozialkompetenzen der Beschäftigten ersetzen. Die Beschäftigten müssen auch unabhängig von KI-Systemen in der Lage sein, ihre Kernaufgaben zu erfüllen. Ihr **Urteilsvermögen bleibt von größter Bedeutung** und geht über die bloße Erfüllung einzelner Aufgaben hinaus. Vor dem Hintergrund des gegenwärtigen Stands der Technik ist insbesondere eine vorschnelle Abhängigkeit notwendiger städtischer Dienstleistungen von Generativer KI zu vermeiden.

(3) Viele KI-Systeme sind interaktiv wie kaum eine Technologie zuvor. Das gilt insbesondere Systeme, die bewusst genutzt werden. KI-Systeme sind keine Wissensdatenbank und auch keine einfachen Automaten, sondern eher ein Kommunikationspartner. **In der Anfangsphase braucht jedes KI-System daher selbst unsere Unterstützung**, damit es uns gezielt unterstützen kann. Das heißt: wir müssen nach und nach lernen, wie ein KI-System uns am besten hilft. Das erfordert Zeit, Erfahrungswerte und Wissen. Schulungsmaßnahmen und Freiräume zum Experimentieren sind in dieser Hinsicht von unerlässlicher Bedeutung.

(4) Aufgrund der rasanten Entwicklung und der kaum vorhandenen Erfahrungen im Alltagsgebrauch von KI-Systemen ist es augenblicklich noch unklar, welchen Mehrwert sie für die Verwaltung liefern und wie ein möglicher Mehrwert sich letztlich ausdrückt. Die zahlreichen Produktivitäts- und Effizienzversprechen, die immer wieder kursieren, entbehren jedenfalls einer tragfähigen empirischen Grundlage. Es handelt sich um Vermutungen, ungewisse Projektionen und Hoffnungen, ohne verlässliche Datenbasis oder Erfahrungswerte. Insofern ist die **Wahrnehmung von KI-Systemen als Unterstützung eine grundlegende Bedingung dafür, dass überhaupt ein Mehrwert entstehen kann**. Mit Unterstützung ist in diesem Grundsatz sowohl eine durch die Menschen erlebte als auch eine durch die Organisation erkennbare und evaluierte Unterstützung gemeint.



(5) **Wenn die Unterstützung von Menschen und Maschinen gelingen soll, darf die Nutzung von KI-Systemen nicht einfach hierarchisch verordnet werden.** Eine nachhaltige Umsetzung erfordert einen kollektiven, partizipativen Aushandlungs- und Lernprozess, der in den jeweiligen Teams gemeinsam mit ihren Führungskräften stattfindet. Dazu braucht es Zeit und Geduld. „KI“ ist weder ein technisches Allheilmittel für die Lösung unklarer Problemlagen der Verwaltung noch kann sie ansatzlos unbestimmte Wünsche nach Effizienzsteigerung erfüllen.



↑ KI muss unsere Zusammenarbeit unterstützen (Quelle: Adobe-Stock)

(6) Das Erleben von KI-Systemen als Unterstützung durch die Beschäftigten ist zugleich **die Grundlage dafür, dass auch Bürger\*innen den Mehrwert dieser Technologie erkennen und davon profitieren**. Eine konkret erlebte Unterstützung der Verwaltungstätigkeiten erleichtert die Bearbeitung bürgerlicher Anliegen und beschleunigt sie im besten Fall. Auf diese Weise nachhaltig eingesetzte KI-Systeme kommen also den Bürger\*innen zugute und können auch nur so von ihnen als Unterstützung wahrgenommen werden. Dass KI-Systeme lediglich unterstützen können, bedeutet auch: Wir bleiben für die Menschen persönlich ansprechbar. So schaffen wir die Grundlagen dafür, KI-Systeme im Sinne der Bürger\*innen einzusetzen.

(7) Wenn ein KI-System zur Unterstützung der eigenen Tätigkeiten genutzt wird, so muss das Ergebnis geprüft werden, bevor es übernommen wird. **Eine ungeprüfte Übernahme der Ausgabe eines KI-Systems widerspricht dem Grundsatz, dass es dienstliche Aufgaben unterstützen soll**. Wird die Ausgabe eines KI-Systems übernommen, gehen wir also davon aus, dass eine menschliche Überprüfung stattgefunden hat. Es liegt dann ein Ergebnis vor, das ein Mensch produziert hat – wenn auch mit Unterstützung durch ein KI-System. Für dieses Ergebnis, seine Verwendung und jegliche Weitergabe ist dann entsprechend die überprüfende und das Ergebnis freigebende Person verantwortlich.

(8) Wird eine dienstliche Aufgabe durch ein KI-System nicht nur unterstützt, sondern ersetzt (z.B. wenn ein Chatbot den Kontakt mit Bürger\*innen übernimmt oder wenn KI-generierte Inhalte unverändert übernommen werden etc.), muss **entsprechend deutlich und für alle sofort wahrnehmbar gekennzeichnet werden, dass man es in der Kommunikation mit einem KI-System zu tun hat**.

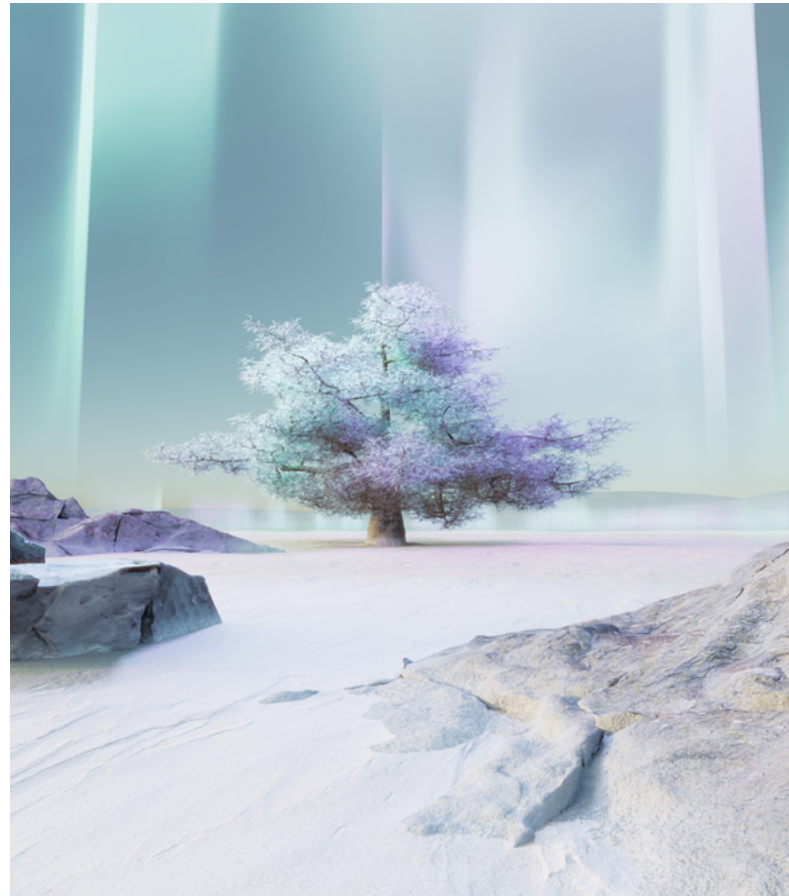


## 3.2 Für Entscheidungen sind immer Menschen verantwortlich

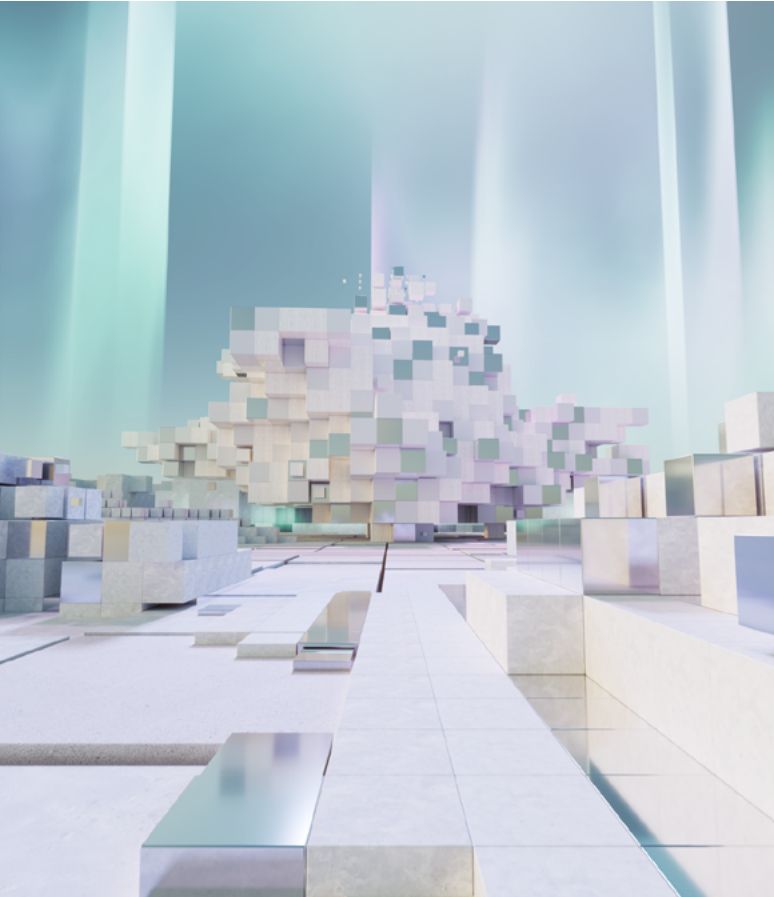
(1) Ein KI-System darf die Entscheidungsfindung unterstützen. Es darf aber nicht autonom entscheiden. Das ist eine zentrale ethische Grenze des Einsatzes von KI-Systemen. **Eine Maschine, die unsere körperliche Erfahrungswelt nicht teilen kann und weder Konsequenzen noch Verantwortlichkeit kennt, darf keine Entscheidungsbefugnis haben.** Es ist darüber hinaus auch eine sehr praktische Grenze. Maschinen entscheiden nicht besser als Menschen, weil das Problem des Entscheidens nicht von der Verfügbarkeit riesiger Datenmengen abhängt, sondern vom Umgang mit Ungewissheit. Die Gleichung „Mehr Daten = Mehr Information = Bessere Entscheidung“ geht in der Praxis nur selten auf.

(2) Die Unterstützung von Entscheidungen durch KI-Systeme kann insbesondere bei unklarer Datenlage und bei interpretationsbedürftigen Daten sinnvoll sein. **Aber ein KI-System kann existierende Formen der Unterstützung und Assistenz nur ergänzen.** Die Möglichkeit einer Unterstützung von menschlichen Entscheidungen durch ein KI-System hat also keinen höheren Stellenwert als andere Formen der organisationalen, kollegialen oder technischen Unterstützung, die jetzt schon genutzt werden, um Informationen für Entscheidungen zu sammeln.

(3) Automatisierung ist nicht identisch mit einem Einsatz von KI-Systemen. Entscheidungen ohne Ermessensspielraum, also gebundene Entscheidungen, können im gesetzlich erlaubten Rahmen automatisiert erfolgen. **Wenn es also darum geht, dass bei Vorliegen derselben Bedingungen immer exakt gleich entschieden werden muss, dann ist der Einsatz von KI-Systemen, insbesondere von Sprachmodellen, überhaupt nicht ratsam.** In diesen Fällen kann ein KI-System die Beschäftigten jedoch beispielsweise dabei unterstützen, das Vorliegen der Bedingungen für eine gebundene Entscheidung festzustellen. Für die Qualität und Richtigkeit der Resultate einer solchen Vorprüfung bleiben weiterhin Menschen verantwortlich.



(4) **Nicht jede Ausgabe oder Aktion eines KI-Systems ist sogleich eine Entscheidung (auch wenn es sich eingebürgert hat, das so zu sagen).** Wenn also auf Grundlage einer KI-gestützten Datenanalyse vordefinierte automatische Aktionen ausgelöst werden, so sind die Entscheidungen schon vorher gefallen. Es sind Entscheidungen, die zum Beispiel festlegen, ob in KI-System eingesetzt werden soll, welche KI eingesetzt werden soll, was sie genau tun soll und welche Aktionen damit ausgelöst werden sollen. Wir achten darauf, dass derartige Eingriffs- und Kontrollmöglichkeiten durch menschliche Entscheidungen immer erfolgen und auch möglich bleiben.



↑ **Wahrnehmung ist etwas anderes als Bildgenerierung**  
(Quelle: Linus Zoll & Google DeepMind [betterimagesofai creative-commons licenses by 4.0](#))

**(5) Betreffen Entscheidungen unmittelbar die Menschen als Bürger\*innen (hoheitliche Entscheidungen), so darf ein KI-System ohne menschliche Überwachung noch nicht einmal vordefinierte Aktionen automatisch auslösen.** Das gilt auch für alle dienstrechtlichen Entscheidungen der Stadtverwaltung als Arbeitgeberin. Werden Menschen durch die Stadtverwaltung hingegen nur als Kund\*innen optionaler Dienstleistungen adressiert, um beispielsweise Angebote zu verbessern, sind solche durch ein KI-System ausgelösten Aktionen durchaus möglich (innerhalb des gesetzlichen Rahmens).

**(6) Wenn es möglich und sinnvoll ist, setzen wir auf eine „Erklärbare KI“ (Explainable AI oder kurz: XAI), um das Mitwirken eines KI-Systems an unseren Entscheidungen besser beurteilen zu können.** XAI befreit uns aber nicht davon, menschliche Verantwortlichkeiten für Entscheidungen festzulegen. „Erklärbarkeit“ ist vereinfacht gesagt dadurch gegeben, dass ein KI-System in Bezug auf die Zuverlässigkeit seines jeweiligen Outputs eine Wahrscheinlichkeit angibt und/oder Teile der verwendeten Daten, auf die es zurückgegriffen hat, rekonstruiert und mit ausgibt. Selbst bei den zu erwartenden, weiteren Fortschritten in diesem Forschungsbereich folgt daraus bis auf Weiteres nicht, dass XAI die Arbeitsweise des verwendeten KI-Systems durchschaubar macht und man damit anschließend genau sagen kann, wie es zu bestimmten Ergebnissen kommt.

**(7) Für den Einsatz und die Nutzung von KI müssen existierende Verantwortlichkeiten nicht zwangsläufig neu bestimmt werden. **Geltende hierarchische und fachliche Verantwortlichkeiten der Mitarbeitenden bleiben auch bei einem Einsatz und einer Nutzung von KI-Systemen bestehen.**** Das gilt auch für die Qualitätssicherung von Entscheidungen und andere Verwaltungsvorgänge.

**(8) Wird der Einsatz eines KI-Systems in Betracht gezogen, das in bestimmten Bereichen tatsächlich autonom entscheiden und agieren soll (das betrifft unter anderem auch den Einsatz von KI-gesteuerten Robotern, vollständig autonom fahrenden Fahrzeugen oder von Agentischer KI), entstehen **ernsthafte Probleme der Zuschreibung von Verantwortlichkeit.**** Daher werden diesbezügliche Entscheidungen, sobald sie in Erwägung gezogen werden, eine gesonderte Einschätzung erforderlich machen.



### 3.3 Wir halten ethische und rechtliche Pflichten ein

**(1) Alle drei Grundsätze dieser Leitlinie adressieren unsere gesellschaftliche, organisationale und ethische Verantwortung.** Sie gründen sich also allesamt auf die Berücksichtigung entsprechender Grundlagen. Dieser dritte Grundsatz betont noch einmal ausdrücklich diese Haltung der Verwaltung und richtet die Aufmerksamkeit darüber hinaus auf die Verknüpfung mit vorhandenen rechtlichen Vorgaben und weiteren ethischen Aspekten.

**(2) Bei einem Einsatz und einer Nutzung von KI-Systemen bleiben alle bereits geltenden Regelungen in Kraft** – unabhängig vom konkreten Einsatzszenario. Das gilt für internationale, nationale, landesspezifische und kommunale Regelungen. Da die Bedeutung dieser technologischen Entwicklung für die Gesellschaft immer wieder betont wird, seien hier explizit und auf höchster Ebene die garantierten Grundrechte genannt. Auf kommunaler Ebene sind es für uns die Charta „Digitales Dortmund“ sowie geltende Dienst- und Geschäftsanweisungen und IT-Standards.

**(3) Wir beachten die KI-Verordnung der EU (kurz: KI-VO oder AI-Act), die dort festgelegten Risikoklassen, die damit verbundenen Verpflichtungen und Transparenzregelungen sowie die Vorgaben zum Kompetenzaufbau.** Dazu gehört auch die Beobachtung der weiteren Entwicklungen der KI-Regulation auf Ebene der EU. Die in der KI-VO festgelegte Klassifikation von KI-Systemen nach ihrem Risiko für das gesellschaftliche Zusammenleben und die Rechte von Individuen dienen uns auch als Orientierung für mögliche Einsatzszenarien.

**(4) Für den Einsatz und die Nutzung von KI-Systemen müssen die datenschutzrechtlichen Vorschriften beachtet werden.** Wir berücksichtigen in diesem Bereich außerdem aktuelle und zukünftige Orientierungshilfen der Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder sowie relevante Erkenntnisse und Empfehlungen anderer datenschutzrechtlicher Instanzen. Darüber hinaus verfolgen wir die aktuelle und zukünftige Rechtsprechung zu KI-Systemen.



**(5) Generative KI-Systeme (z.B. ChatGPT, Copilot, Gemini etc.) werden von uns grundsätzlich wie öffentliche Räume behandelt.** Personenbezogene Daten oder vertrauliche Informationen dürfen dort also nicht eingegeben bzw. hochgeladen werden. Ein Grund dafür: Diese hochsensiblen Daten können trotz bestimmter Sicherheitsvorkehrungen unter Umständen in den Ergebnissen von Anfragen anderer Nutzer\*innen wieder auftauchen.

**(6) Wir berücksichtigen existierende ethische Richtlinien und den Stand wissenschaftlicher Forschung zu KI-Technologien** – zum Beispiel in den Bereichen der Science and Technology Studies, der Mensch-Technik-Interaktion, der Technikfolgenabschätzung und der Technikethik.



↑ Wir müssen KI-Systeme mit unseren Daten füttern – und doch funktionieren sie ganz anders als wir (Quelle: Yutong Liu & Kingston School of Art [betterimagesofai creativecommons licenses by 4.0](#))

**(7) Im Fall der zukünftigen Entwicklung eigener KI-Lösungen berücksichtigen wir den Ansatz „Ethics by Design“.**<sup>2</sup> Das heißt, dass wir von Beginn an besondere Sorgfalt in Bezug auf zwischenmenschliche, organisationale, gesellschaftliche und damit ver-

knüpfte ethische Fragen walten lassen.

**(8) Nutzen wir KI-Systeme und/oder KI-Modelle externer Anbieter, so achten wir in einem zumutbaren Aufwand darauf, dass sie sich zur Einhaltung bestimmter ethischer Standards verpflichtet haben.** Das ist leider kaum zuverlässig überprüfbar. **Deshalb sorgen Maßnahmen zum Aufbau von KI-Kompetenzen dafür, dass die Beschäftigten die Qualität und Nutzbarkeit möglicher Outputs eines KI-Systems beurteilen können.** Beschäftigte müssen für Urheberrechtsverletzungen sensibilisiert werden und außerdem lernen, diskriminierende Inhalte zu erkennen und ihre Verbreitung zu verhindern.

**(9) Der Betrieb von Modellen Generativer KI, insbesondere besonders großer und leistungsfähiger Sprachmodelle (Large Language Models, LLMs), verbraucht sehr viel Energie – mehr Energie als andere alltägliche Aktivitäten im Internet. Es ist zwar zu erwarten, dass sich die Energieeffizienz von KI-Modellen in Zukunft nach und nach verbessert. Der Betrieb fortgeschrittener KI-Modelle wird insgesamt jedoch energie- und ressourcenintensiv bleiben.** Sofern der Einsatz oder die Nutzung von Generativer KI in der Verwaltung also intensiviert wird, trägt die Stadt ihren Teil dazu bei, den weltweiten Energieverbrauch zu erhöhen. Daher ist es in Bezug auf unsere ökologische Verantwortung und im Sinne der Nachhaltigkeit notwendig, in unseren Entscheidungen auch stets abzuwägen und zu evaluieren, in welchem Ausmaß ein Einsatz von KI-Systemen vor diesem Hintergrund sinnvoll ist. Das gilt auch für die alltägliche Nutzung insbesondere von Generativer KI – von der Sachbearbeitung bis hin zur Führung.

**(10) Für eine verantwortliche Gestaltung von KI-Systemen im kommunalen Bereich müssen wir laufend auch die Möglichkeiten eines Aufbaus von Entwicklungskompetenzen erkunden.** Unter Umständen müssen wir selbst in der Lage sein, kleine und auf unsere Bedarfe zugeschnittene KI-Modelle zu bauen. Solche Modelle verbrauchen nicht nur weniger Energie, sondern können auch unsere digitale Souveränität stärken. Im besten Fall kann und soll das auch im Verbund mit anderen Kommunen geschehen.

<sup>1</sup> Dazu gehören unter anderem die [Ethik-Leitlinien für eine vertrauenswürdige KI der EU](#) oder auch das [Gutachten der Datenethikkommission der Bundesregierung](#)

<sup>2</sup> Hinweise zu diesem Ansatz finden sich zum Beispiel in einem entsprechenden [Dokument der Europäischen Kommission](#)



An aerial night photograph of a city, likely Utrecht, showing a dense urban landscape with numerous illuminated buildings. A prominent building in the center features a large, brightly lit 'U' logo on its roof. The sky is a mix of deep blue and orange from the setting or rising sun. The foreground is dominated by a large purple rounded rectangle containing text.

# **Zum Schluss**

Auf wen und worauf  
gründet sich diese Leitlinie?

**Diese Leitlinie gründet sich auf die Arbeit und die Diskussionen einer fachbereichsübergreifenden Arbeitsgruppe der Stadtverwaltung Dortmund zum Thema Künstliche Intelligenz (KI-DO).**

Dort treffen sich Expert\*innen aus verschiedenen Fachbereichen (u.a. Dortmunder Systemhaus sowie Personal- und Organisationsamt) mit der Datenschutzbeauftragten, der Compliance-Beauftragten und dem Personalrat, um über Bedingungen und Möglichkeiten des Einsatzes von KI-Systemen in Dortmund zu diskutieren und zu beraten.

Die Arbeitsgruppe wird durch PD Dr. Athanasios Karafilidis koordiniert und fachlich geleitet. Er ist KI-Manager der Stadt Dortmund und hat auf Grundlage der Diskussionen in der Arbeitsgruppe, seines Austauschs mit Expert\*innen und Beauftragten anderer Kommunen und Institutionen sowie unter Berücksichtigung des Stands der fachlichen Diskussionen (in der Technik- und Organisationsforschung und zum Einsatz von KI-Systemen in der öffentlichen Verwaltung) diese Leitlinie formuliert und in mehreren Feedbackrunden überarbeitet.

**Die Leitlinie ist am 18.02.2024 in den Verwaltungsvorstand eingebracht und dort diskutiert und beschlossen worden.**



Einige wissenschaftliche Grundlagen für die Aussagen in dieser Leitlinie finden sich unter anderem in folgenden Publikationen (Auswahl):

**Besio, Cristina/Fedtke, Cornelia/Grothe-Hammer, Michael/Karafilidis, Athanasios/Pronzini, Andrea (2024).** Algorithmic Responsibility Without Accountability: Understanding Data-Intensive Algorithms and Decisions in Organisations. *Systems Research and Behavioral Science* 2024: 1-17.

**Bischof, Andreas (2017).** Soziale Maschinen bauen. Epistemische Praktiken der Sozialrobotik. Bielefeld: transcript.

**Collins, Harry (2018).** Artificial Intelligence. Against Humanity's Surrender to Computers. Cambridge: polity.

**Esposito, Elena (2024).** Kommunikation mit unverständlichen Maschinen. Wien/Salzburg: Residenz.

**Hirsch-Kreinsen, Hartmut/Krokowski, Thorben (2024).** Promises and Myths of Artificial Intelligence. *Weizenbaum Journal of the Digital Society* 4 (1): 1-8.

**Karafilidis, Athanasios (2019).** Unterstützung und Assistenz durch die Maschine. In: Kevin Liggieri und Oliver Müller (Hrsg.): *Mensch-Maschine-Interaktion. Handbuch zur Geschichte – Kultur – Ethik.* Stuttgart: J.B. Metzler, S. 157-164.

**Karafilidis, Athanasios (2020).** *Hybride Interfaces. Elemente einer soziologischen Theorie technischer Unterstützung.* Habilitationsschrift, RWTH Aachen.

**Luhmann, Niklas (1966).** *Recht und Automation in der öffentlichen Verwaltung. Eine verwaltungswissenschaftliche Untersuchung.* Berlin: Duncker & Humblot.

**Luhmann, Niklas (2000).** *Organisation und Entscheidung, Opladen/Wiesbaden: Westdeutscher Verlag.*

**Marcus, Gary (2024).** *Taming Silicon Valley. How We Can Ensure that AI Works For Us.* Cambridge: MIT Press.

**Narayanan, Aervind/Kapoor, Sayash (2024).** *AI Snake Oil. What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference.* Princeton: Princeton UP.

**Zweig, Katharina (2023).** Droht KI den Menschen zu ersetzen? *Aus Politik und Zeitgeschichte* 73 (42): 4-8.

### Weitere Bildquellen:

**Cover:** Yutong Liu & Digit  
[betterimagesofai creativecommons licenses by 4.0](#)

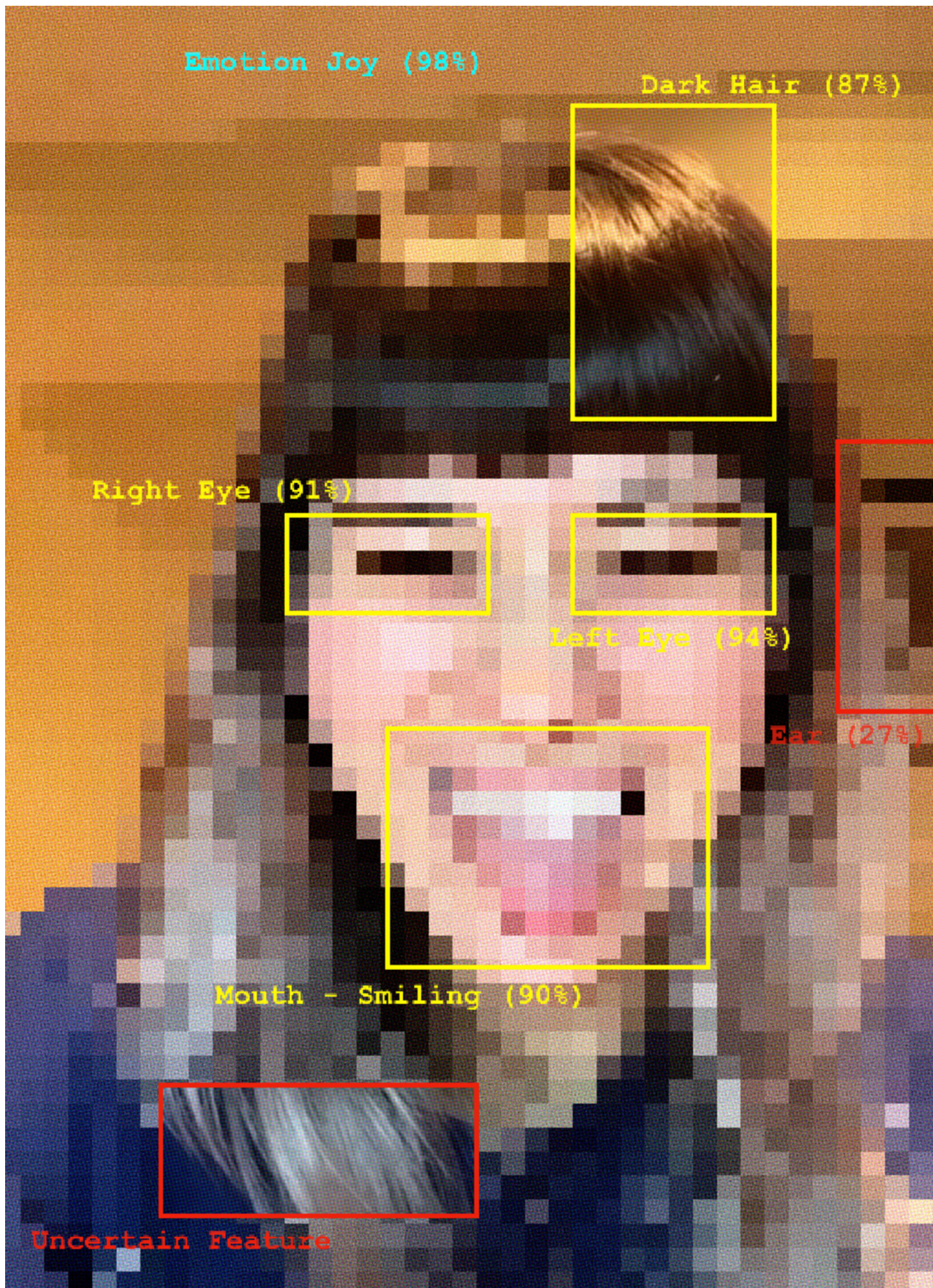
**Seite 2:** Roland Gorecki

### Zum Ende ein wichtiger Hinweis:

Die Vielfalt möglicher Einsatzbereiche von KI in der Verwaltung, die Dynamik der Entwicklung von KI und die generelle Unmöglichkeit, alle denkbaren Szenarien zu berücksichtigen macht es erforderlich, dass dieses Dokument, ebenso wie andere damit verknüpfte Regelungen (die „Geschäftsweisung KI“ und das „Handbuch KI“) regelmäßig auf ihre Aktualität und Angemessenheit überprüft und falls notwendig entsprechend angepasst werden.

Letzte Aktualisierung im Juni 2026







## **Impressum**

**Herausgeberin** Stadt Dortmund – Dortmunder Systemhaus **Redaktion und Bildauswahl**

Athanasios Karafillidis (verantwortlich) **Bilder** Stadt Dortmund – Roland Gorecki, Adobe

Fotostock, Better Images of AI **Konzept, Design** Stadt Dortmund, Marketing + Kommunikation in  
Zusammenarbeit mit bandefürgestaltung 07/2026