

## MA RWP Lectures 5

Peter G. Kirchschräger

# ***Digitalisierung und «künstliche Intelligenz» – Maschinen oder Menschen am Steuer?***

Vernetzt  
forschen.

Verantwortlich  
handeln.



# MA RWP Lectures

## Band 5



Peter G. Kirchschräger

Digitalisierung und «künstliche  
Intelligenz» – Maschinen oder  
Menschen am Steuer?

Mit einem Vorwort von  
Antonius Liedhegener

Herausgegeben von  
ZRWP / Universität Luzern

Luzern 2024



## Vorwort

Die Digitalisierung gehört zu den grossen Motoren der gesellschaftlichen Veränderung in der Gegenwart. Vergleichbar der industriellen Revolution im 19. Jahrhundert führt sie zu einer umfassenden Veränderung der sozialen Beziehungen und gesellschaftlichen Strukturen. Am deutlichsten zu spüren ist dies in der Veränderung der Öffentlichkeit und der Arbeitswelt. Die von Algorithmen getriebene Kommunikation in den Social Media haben den klassischen Medienbereich und -markt umgewälzt, mit weitreichenden Folgen für eine Politik im Echtzeitmodus des Internets und für das Vertrauen der Bürgerinnen und Bürger in Medien und Politik. Die Abbildung von realen Produktions- und Distributionsprozessen und -ketten in digitaler Form, die intelligente Produktion durch Roboter und die Entstehung einer E-Commerce-Kultur schichten Arbeitsmärkte und Landschaften um, viele Berufsgruppen und urbane Innenstadträume sehen sich in ihrer Existenz bedroht.

Die Innovationsschübe sind schnell, weitreichend und nachhaltig. Zur Erinnerung: Nach langem Vorlauf in Wissenschaft und Militär wird das Internet um 1990 allgemein nutzbar. 1993 geht das Worldwide Web an den Start. E-Mails werden zu einem

Standardwerkzeug der Kommunikation in modernen Gesellschaften. Mit Videoplattformen entstehen neue Kommunikationswege und -formen. Durch die Sozialen Netzwerk wird das Internet Ende der 1990er Jahre dialogisch. Mobilfunknetze verändern und erweitern Räume und Raumwahrnehmungen. Smartphones bringen die neuen Möglichkeiten seit 2007 nahezu weltweit in die Hand und in den Alltag der Userinnen und User. Und aktuell sticht der Durchbruch der Künstlichen Intelligenz (KI) hervor. Wenige Innovationen haben sich so schnell in den Alltag der Menschen eingeschrieben wie etwa der Chatbot ChatGPT oder KI-basierte Programme zur Bild- und Musikbearbeitung. Sie verändern die sozialen Bedingungen der Wissens- und Medienproduktion anscheinend fundamental. Insbesondere werden Information und Manipulation, werden Realität und Virtualität, werden Wahrheit und Unwahrheit vielfach ununterscheidbar. Die Echtheit eines Fotos oder Bildes, das auf dem Smartphone- oder PC-Bildschirm erscheint, zu bestimmen, ist oft zu einer nicht lösbaren Aufgabe geworden. Existiert diese schöne Landschaft, die mir gezeigt wird, irgendwo auf der Welt in natura?



Ob der Durchbruch der KI am Massenmarkt einen gleichen Stellenwert einnehmen wird wie die Erfindung des Buchdrucks oder das Einsetzen der industriellen Revolution, ist eine Frage, die sich aktuell wohl noch nicht sicher beantworten lässt. Dass die Folgen aber noch weit über die eingangs angesprochenen Phänomene hinausreichen, zeigt der Sozialethiker Peter G. Kirchschräger in der vorliegenden fünften MA RWP Lecture auf. Chancen wie Gefährdungen der als Künstliche Intelligenz bezeichneten Computertechnologien stellen die Entscheidungs- und Handlungsfähigkeiten der Einzelnen wie der Menschheit insgesamt vor grosse Herausforderungen. Pointiert gefragt: Beendet das autonome Fahren die Autonomie menschlichen Handelns? Übernimmt die Technik das Steuer, wenn es um Entscheidung, Verantwortung und moralisches Urteilen geht? Peter G. Kirchschräger ist überzeugt, dass eine menschliche Zukunft nur im Rahmen einer internationalen Regulierung gefunden werden kann. Seine Lecture zeigt die Gründe dafür auf und endet mit einem ganz konkreten Vorschlag, wie auf der Basis der Menschenrechte die Einhegung der KI-Technologien durch internationale Standards, die definiert, überprüfbar und durchsetzbar sind, gelingen sollte. Selbstverständlich bleiben auch das individuelle Handeln und der

Umgang mit KI wichtig, aber aus ethischer Perspektive ist die anstehende Aufgabe für ihn vor allem eine institutionenethische Frage.

Die fünfte MA RWP Lecture kann und soll dazu anregen, die Frage nach dem richtigen, ethisch vertretbaren Umgang mit den neuen Technologien zu durchdenken und diskutieren. Die Studierenden des Joint Degree Masterstudiengangs Religion, Wirtschaft, Politik haben das Thema und den Referenten der fünften MA RWP Lecture vorgeschlagen. Es freut uns, dass Peter G. Kirchschräger sich bereit erklärt hat, seine Ideen in die Lecture und die vorliegende Publikation einzubringen. Wir danken ihm für seine umfassende Schilderung der gegenwärtigen Lage, die zahlreichen Einsichten aus einer globalen Perspektive, die sozialetischen Überlegungen und den Lösungsvorschlag für eine internationale Regulierung analog zu anderen Menschheitsthemen.

Die Reihe der MA RWP Lectures veröffentlicht in loser Folge Beiträge, die Denkangebote und neue wissenschaftliche Erkenntnisse auf den Schnittfeldern von Religion, Wirtschaft und Politik präsentieren. Getragen wird die Reihe vom Joint Degree Master Religion, Wirtschaft und Politik. Diese Aus-

gabe hat Frau Dr. Silvia Martens in ihrer Eigenschaft als Koordinatorin des MA RWP an der Universität Luzern in bewährter Form redaktionell betreut. Ihr gilt ein herzliches Dankeschön dafür. Ein grosses Dankeschön geht an den Kollegen Peter G. Kirchschräger für seine Bereitschaft, die fünfte MA RWP Lectures zu halten. Den Text seiner Vorlesung legen wir hiermit der Öffentlichkeit vor. Ich wünsche dem Text eine weite Verbreitung und den Leserinnen und Lesern durch die Lektüre neue Einsichten und einen Gewinn an ethischer Analyse und Bewertung jener Technologien, die wir nunmehr KI nennen.

Prof. Dr. Antonius Liedhegener

Vorsitzender der Studiengangsleitung des MA RWP

Zentrum für Religion, Wirtschaft und Politik  
(ZRWP), Universität Luzern

## Zum Autor



### **Prof. Dr. theol. lic. phil. Peter G. Kirchschräger**

Peter G. Kirchschräger ist seit 2017 Ordinarius für Theologische Ethik und Leiter des Instituts für Sozialethik (ISE) der Theologischen Fakultät der Universität Luzern. Er ist Gastprofessor an der ETH Zürich und beratender Experte in ethischen Fragen für nationale und internationale Organisationen und Institutionen (z. B. für die UN, UNESCO, die Organisation für Sicherheit und Zusammenarbeit in Europa [OSZE], die Europäische Union [EU], den Europarat, Unternehmen und NGOs).

## Digitalisierung und «künstliche Intelligenz» halten die Welt in Bann

Unternehmen wälzen mit Angeboten von Digitalisierung und des Einsatzes von sogenannter «künstlicher Intelligenz (KI)» das Leben der Menschen privat und beruflich um. Beispielsweise wird jede Onlinerecherche – der wesentliche Informations- und Wissenszugang der Gegenwart – durch sogenannte «KI» individuell «optimiert». Grosse Sprach- und Bildergenerierungsmodelle (wie z. B. ChatGPT, Dall-E) produzieren semantisch und syntaktisch korrekte Sätze bzw. ansprechende Visualisierungen und entlasten dadurch Menschen in der Wahrnehmung ihrer beruflichen Aufgaben und in ihrer persönlichen Kommunikation.

Aus ethischer Perspektive ist angesichts und im Zuge des umfassenden technologiebasierten Wandels nach den ethischen Chancen und den ethischen Risiken der Digitalisierung und der sogenannten «KI» zu fragen (Završnik & Simončič, 2023; Boddington, 2023; Risse, 2023). Das ethisch Positive und Negative ist zu identifizieren, um die ethischen Sonnenseiten der Digitalisierung und der sogenannten «KI» fördern und die ethischen Schattenseiten vermeiden oder meistern zu können (Kirchschräger, 2024a). Damit verbunden stellt sich

auch die Frage nach der Verantwortung: Soll angesichts der Tatsache, dass sogenannte «KI» den Menschen in gewissen Aufgaben und Bereichen bereits massiv überträgt, das Steuer der «KI» überlassen werden? Was sollen Menschen mit den Möglichkeiten digitaler Transformation und «KI» tun? Soll alles technisch Machbare auch realisiert werden?

In einem ersten Schritt dieser ethischen Betrachtung gilt es, die Begriffe «Digitalisierung» und «künstliche Intelligenz» (KI) aus ethischer Sicht zu beleuchten, was zum Vorschlag einer genaueren Bezeichnung der sogenannten «KI» als «datenbasierte Systeme» (im Folgenden kurz DS) führt.

Damit verbunden ist der zweite Schritt der Frage gewidmet, ob – und falls ja – in welchen Bereichen die Menschen der Digitalisierung und DS das Feld überlassen sollen und wo den Menschen eine Mit- oder Alleinverantwortung zukommen soll.

Der dritte Schritt führt die Menschenrechte als ethischen Referenzrahmen ein, der in der Folge als Maßstab für die ethische Bewertung von Digitalisierung und DS dient.

Der vierte Schritt fokussiert auf die Identifizierung von ethischen Potentialen und Gefahren von Digitalisierung und DS anhand des ethischen Referenzrahmens der Menschenrechte.

In einem fünften Schritt soll dann auf konkrete Möglichkeiten der Förderung der ethischen Chancen und der Vermeidung bzw. Meisterung von ethischen Risiken von Digitalisierung und DS eingegangen werden. Dabei stehen insbesondere zwei konkrete Vorschläge im Fokus, die bereits weltweit ein positives Echo und Unterstützung ausgelöst haben: menschenrechtsbasierte DS und die Schaffung einer Internationalen Agentur für datenbasierte Systeme (IDA) bei der UNO.

### **Digitalisierung und datenbasierte Systeme (DS)**

Der Begriff «Digitalisierung» steht für verschiedene technologiebasierte Veränderungen, die «Digitales» umfassen. «Digital» bedeutet «Aufzeichnung oder Speicherung von Informationen als eine Reihe der Zahlen 1 und 0, um zu zeigen, dass ein Signal vorhanden ist oder nicht, unter Verwendung von oder im Zusammenhang mit digitalen Signalen und Computertechnologie, Darstellung von Informationen in Form eines elektronischen Bildes, unter

Verwendung eines Systems, das von einem Computer und anderen elektronischen Geräten verwendet werden kann, in dem Informationen in elektronischer Form als eine Reihe der Zahlen 1 und 0 gesendet und empfangen werden, im Zusammenhang mit Computertechnologie, insbesondere dem Internet, Darstellung von Informationen als ganze Zahlen und nicht in einer anderen Form wie einem Bild, einer Grafik usw.» (Cambridge Dictionary, o. D.). Dabei ist zu berücksichtigen, dass Objekte, Phänomene und Realitäten als Digitalisierung bezeichnet werden, obwohl sie selbst nicht digital sind, sondern Produkte oder Ergebnisse digitaler Prozesse darstellen (Ohly, 2019, S. 25-29).

Sogenannte «künstliche Intelligenz (KI)» kann definiert werden als Streben mit technischen Mitteln nach der Nachahmung oder Erfüllung von kognitiven Funktionen des menschlichen Denkens. Angesichts des Wesens der «KI» sind aus ethischer Perspektive Zweifel angebracht, ob der Begriff angemessen ist, denn «KI» verfolgt das Ziel, die menschliche Intelligenz zu imitieren, was aber auf einen bestimmten Bereich der Intelligenz (z.B. bestimmte kognitive Fähigkeiten) beschränkt ist (Dreyfus, 1972; Dreyfus & Dreyfus, 1986). Weiterhin ist davon auszugehen, dass «KI» allenfalls in



bestimmten Bereichen der Intelligenz der menschlichen Intelligenz ähnlich werden und diese simulieren kann, aber niemals gleich werden kann. Unter anderem im Bereich der emotionalen und sozialen Intelligenz sind Maschinen nur in der Lage, Emotionen, persönliche Interaktion und Beziehungen nachzuahmen, und es fehlt ihnen an Authentizität. Einem Pflegeroboter kann beispielsweise beigebracht werden, zu weinen, wenn Patient:innen weinen, aber niemand würde behaupten, dass der Roboter echte Gefühle empfindet und deshalb weint. Umgekehrt könnte man denselben Roboter darauf trainieren, Patient:innen eine Ohrfeige zu geben, wenn diese weinen, und der Roboter würde diese Funktion auf die gleiche konsequente Weise ausführen.

Darüber hinaus kann man Maschinen keine Moralfähigkeit zuschreiben, weil sie nur von Menschen vorgegebenen ethischen Prinzipien und Normen folgen; sie können nicht selbst ethische Regeln erkennen und setzen. Technologien werden in erster Linie nach ihrer Tauglichkeit hergestellt und können als selbstlernende Systeme z. B. effizienzsteigernde Regeln aufstellen. So könnte z. B. ein selbstfahrendes Auto sich selbst die Regel geben,

Menschen und die Natur zu schädigen, um möglichst schnell von A nach B zu kommen, ohne dass es in der Lage wäre, ethische Regeln für sich selbst zu erkennen, die es der Maschine erlauben würden, die Illegitimität dieser Vorgabe zu erkennen. Eine menschliche Person am Steuer hingegen besitzt das Potenzial, für sich selbst verbindliche ethische Regeln zu erkennen, die sie zur Einsicht befähigen, dass die Schädigung von Menschen und Natur zwar effizienter, aber illegitim ist. Maschinen fehlt diese Autonomie. Zur Autonomie gehört es, ethische Normen für sich selbst zu erkennen, zu setzen und das eigene Handeln daran auszurichten (Kant, 1974). Maschinen scheitern am Prinzip der Verallgemeinerbarkeit. Dieses Prinzip setzt voraus, dass rationale und plausible Argumente – «gute Gründe» – vorgebracht werden. «Gute Gründe» bedeutet, dass es denkbar sein muss, dass alle Menschen in ihrer effektiven Freiheit und Autonomie sowie ihrer vollen Gleichheit diesen Gründen – innerhalb eines Denkmodells und nicht innerhalb eines realen weltweiten Referendums – aus ethischen Gründen zustimmen würden (Kirchschräger, 2023).

Als Alternative zur Rede von sogenannter «künstlicher Intelligenz» wäre der Begriff «datenbasierte Systeme (DS)» angemessener (Kirchschräger, 2021),

weil dieser Begriff das beschreibt, was diese Technologien eigentlich ausmacht: Generierung, Sammlung und Auswertung von Daten, datenbasierte Wahrnehmung (sensorisch, sprachlich), datenbasierte Vorhersagen, datenbasierte Entscheidungen. Darüber hinaus erlaubt der Begriff «datenbasierte Systeme», die wichtigsten Stärken und Schwächen der derzeitigen technologischen Errungenschaften in diesem Bereich hervorzuheben. Der Hinweis auf ihre Kerneigenschaft, nämlich auf Daten zu beruhen und sich in all ihren Prozessen, ihrer eigenen Entwicklung und ihren Handlungen – genauer gesagt, ihren Reaktionen auf Daten – ausschliesslich auf Daten zu stützen, lüftet den Schleier der unangemessenen Zuschreibung des Mythos «Intelligenz» und legt wesentliche ethische Probleme und Herausforderungen von DS frei (Kirchschräger, 2022). Dies ermöglicht mehr Genauigkeit, Angemessenheit und Präzision bei der kritischen Reflexion über DS und unterbindet ein ethisch problematisches Übervertrauen in DS.

### **Ethische Entscheidungen als exklusive Verantwortung der Menschen**

Die bisherigen Überlegungen legen nahe, dass aufgrund der fehlenden Moralfähigkeit von Maschinen

den Menschen die exklusive Verantwortung zukommt, ethische Entscheidungen zu fällen. DS können wegen ihres oben dargelegten Mangels an Autonomie nicht als moralische Subjekte auftreten. Auch die Festlegung ethischer und rechtlicher Prinzipien und Normen sowie von einem Rahmen für die Digitalisierung und DS gehören in die Hände von Menschen. Denn das Potenzial, über das Technologien in Bezug auf ethische Entscheidungen und Handlungen verfügen, reicht bei weitem nicht an die Moralfähigkeit von Menschen heran, da es Maschinen nicht nur an Autonomie, sondern auch an Verletzbarkeit, Gewissen, Freiheit und Verantwortung fehlt, die für Moralfähigkeit unerlässlich sind (Kirchschräger, 2021). Ebenfalls tragen Menschen die Verantwortung dafür, technologiebasierte Innovationen unter ethischen Gesichtspunkten zu prüfen, zu analysieren und zu bewerten sowie nicht alles technisch Machbare zu verwirklichen. Dabei gilt es auch, der Komplexität der Ethik Rechnung zu tragen, zu der u.a. die «dual use»-Problematik, die Ambivalenz und die regelüberragende Einzigartigkeit des Konkreten zählen (Kirchschräger, 2023). Letztere hebt hervor, dass in einer konkreten Begegnung mit konkreten Menschen in einer konkreten Situation ethische Regeln an ihre Grenzen stoßen können, weil das Konkrete in

seiner Einzigartigkeit die Regel überragt. Durch die zunehmende Komplexität der Alltagswirklichkeit – z. B. wenn ethische Prinzipien divergieren oder kollidieren – ist der Mensch herausgefordert, das ethisch Richtige in einer konkreten Begegnung mit konkreten Menschen in einer konkreten Situation herauszufinden. Aufgrund ihrer fehlenden Moralfähigkeit stoßen DS hier an Grenzen. Eine vollständige Übertragung der Ethik auf die Mathematik, die Programmierung oder das Training von Maschinen erweist sich aufgrund der fehlenden Moralfähigkeit von DS sowie wegen Komplexität der Ethik als unmöglich.

Die ethische Verantwortung von Menschen für DS bedeutet auch Gestaltungsverantwortung: Digitalisierung und DS fallen nicht einfach vom Himmel oder sind von der Natur vorgegeben, sondern es ist an den Menschen zu entscheiden, was und wie Digitalisierung und DS gestaltet, entwickelt, hergestellt, eingesetzt und verwendet bzw. auch aus ethischen Gründen nicht genutzt werden. Zum Letzteren gehört auch die Verantwortung, Digitalisierung und DS dort gezielt zum Zuge kommen zu lassen, wo sie der Menschheit und dem Planeten dienen, sowie DS in den Bereichen menschlicher

Intelligenz zur Entfaltung zu bringen, wo sie die Menschen bereits massiv überragen – wie z. B. Umgang mit umfangreichen Datenmengen, Datenspeicherung, Rechenfähigkeit, logische Deduktion. DS können das menschliche Denken durch Faktenspeicherung und Komplexitätswahrnehmung substantiell unterstützen.

## **Menschenrechte als ethischer Referenzrahmen für Digitalisierung und DS**

Um diese ethische Verantwortung für Digitalisierung und DS wahrzunehmen, gilt es, Digitalisierung und DS von einem ethischen Standpunkt zu prüfen. Auf der Suche nach einem ethischen Referenzrahmen für diese ethische Analyse liegt es nahe, mit einem ethischen Minimalstandard zu beginnen, um den technologischen Fortschritt nicht ethisch zu überfordern und mit einem hohen Ethos zu erdrücken. Gleichzeitig sollte die ethische Orientierungshilfe globale Akzeptanz aufweisen.

Die Menschenrechte könnten als ethischer Bezugsrahmen als Mindestanforderung die notwendige normative Orientierung bieten, dass sie nur einen Mindeststandard darstellen, der allen Menschen immer und überall ein physisches Überleben und ein Leben in Würde – ein lebenswertes Leben –

garantiert. Für die Menschenrechte spricht zudem die ethische Begründbarkeit der Menschenrechte und ihrer Universalität mit dem Prinzip der Verletzbarkeit (Kirchschräger, 2013a), was ihre globale Akzeptanz untermauert. Letztere erweist sich als konkurrenzlos im Vergleich zu anderen Wertesystemen und Normenkatalogen (Kirchschräger, 2016a), da die Menschenrechte nicht in einer bestimmten Kultur, Tradition, Religion oder Weltanschauung verankert sind, sondern als Basis für die gelebte Vielfalt von Kulturen, Traditionen, Religionen oder Weltanschauungen dienen, indem sie u.a. die Glaubens-, Gewissens- und Religionsfreiheit aller Menschen schützen.

Ausserdem ermutigen und fördern die Menschenrechte Innovationen und technologischen Fortschritt, indem sie die Gedankenfreiheit, die Meinungsäusserungsfreiheit und den freien Zugang zu Informationen schützen (Kirchschräger, 2013b). Schliesslich schützen und fördern sie die für Innovationsprozesse konstitutive Vielfalt durch die Garantie der Selbstbestimmung jedes Menschen, was sich als Quelle von Vielfalt erweist (Kirchschräger, 2016a; 2020).

## Menschenrechtspotential und Menschenrechtsverletzungen von DS

Digitalisierung und DS können einen wichtigen Beitrag zur Förderung von Menschenrechten leisten. Beispielsweise werden mit DS neue Medikamente anhand von Proteinstrukturen entworfen (Atz et al., 2024). Tumorerkrankungen können mit Hilfe von DS präziser als bisher hinsichtlich ihrer Bedrohlichkeit eingeordnet werden (Zhang et al., 2024). DS bringen enorme ethische Chancen mit sich.

In der Beschäftigung mit den ethischen Chancen und Risiken von DS werden im Fachdiskurs dazu allzu oft die folgenden beiden Denkfehler gemacht, die vermieden werden sollten. Erstens: Ethische Gefahren könnten mit ethischen Gelegenheiten vermeintlich ausgeglichen werden. Zweitens: Menschenrechtsverletzungen werden in Kauf genommen, wenn ein grosser Nutzen von DS erwartet wird. Dabei handelt es sich um zwei Denkfehler, weil eine Verletzung von Menschenrechten mit oder durch DS weder durch das positive ethische Potential von DS im Sinne der Menschenrechte noch durch einen hohen Ertrag aufgewogen werden kann, da es um Menschenrechte geht, sodass sich



jede einzelne Verletzung von Menschenrechten als ethisch inakzeptabel erweist.

Umso mehr fallen die ethischen Risiken ins Gewicht. Aus ethischer Perspektive besteht die erste globale Gefahr von DS darin, nichts gegen die ethischen Risiken von DS zu unternehmen: Produkte mit DS, die die Menschenwürde und die Menschenrechte verletzen oder den Planeten zerstören, können ohne rechtliche Konsequenzen auf den Markt gebracht werden. Wenn beispielsweise eine App online verkauft wird, die Bilder von Kindern sexualisiert, passiert dem Unternehmen, das diese App anbietet, nichts ausser, dass es viel Geld verdient (Lensa AI, o.D; Heikkilä, 2022; Snow, 2022). Mehrere Jahrzehnte hat die Menschheit nun Geschäftsmodelle unbehelligt zugelassen, die in ihrem Kern Menschenrechtsverletzungen beinhalten.

Eine zweite globale Gefahr durch DS ist die wachsende Ungleichheit und Armut auf der ganzen Welt, die auf eine immer grösser werdende «digitale Kluft» zurückzuführen ist (Kirchschräger, 2021). Zu erwartende explosionsartige Steigerungen der Recheneffizienz (z. B. durch Quantum-Computing) wird die bereits bestehende «digitale Kluft» – in einigen

Regionen der Welt gibt es keinen Zugang zum Internet, andere Orte profitieren bereits von grossen Rechenzentren – explodieren lassen.

Eine dritte globale Gefahr stellen die negativen Auswirkungen von DS auf das Klima und die Umwelt dar, die den Planeten zu zerstören drohen (Kirchschräger, 2021). Der massive Bedarf sowohl an Rohstoffen zur Produktion von DS als auch an Energie bei der Herstellung und beim Einsatz von DS steht noch in keinem Verhältnis zum Beitrag von DS zur Bekämpfung der Klimazerstörung.

Eine vierte globale Gefahr umfasst die ständige Verletzung des Menschenrechts auf Privatsphäre und Datenschutz (Zuboff, 2019). Wann immer möglich, werden Daten von Menschen gestohlen und an den Meistbietenden verkauft. Die anhaltende Missachtung der Privatsphäre und des Datenschutzes ist ein massiver Angriff auf die Freiheit aller Menschen.

Eine fünfte globale Gefahr besteht darin, dass man mit DS dank dieser riesigen Datenmengen über Menschen die Menschen besser «kennt» als die Menschen sich selbst. Dies öffnet die Tür für wirtschaftliche und politische Manipulation. Mit DS kann man genau wissen – um eine Metapher zu

verwenden –, welche Klaviertasten gedrückt werden müssen, um die Musik zum Klingen zu bringen, mit anderen Worten, um die Menschen dazu zu bringen, so zu kaufen oder zu wählen bzw. abzustimmen, wie es die die DS kontrollierenden Hände wollen (Nosthoff & Maschewski, 2019).

Eine sechste globale Gefahr sind die Potentiale der Desinformation. Aufgrund der globalen Vernetzung online können mit einem Mausklick Millionen von Menschen gezielt mit falschen Informationen – sogenannten «Fake News» – eingedeckt werden. Sprach- und Bildgenerierungssysteme (wie z. B. ChatGPT, Dall-E), für die das Kriterium der Wahrscheinlichkeit, Datenschutz und Urheberrechte keine Rollen spielen und die bildlich gesprochen nichts Anderes als wiederkäuende Kühe sind (Kirchschräger, 2024b), eröffnen diesbezüglich neue Horizonte, da künstlich mit hohem Perfektionsgrad geschaffenes Audio- und Videomaterial als vermeintliche Realitäten global in Umlauf gebracht werden können. So können beispielsweise politischen Amtsträger:innen Worte in den Mund gelegt werden, die sie nie so gesagt haben. Leider bleibt trotz offizieller Klarstellungen seitens der betroffenen Politiker:innen etwas von den Unwahrheiten in den Köpfen der Menschen hängen – bei «Fake

News» in Bildern noch mehr als in Worten (Kirchschräger, 2021).

Damit verbunden erweist sich als eine siebte globale Gefahr die Möglichkeit der gezielten Destabilisierung von Gesellschaften durch auf dem Internet – insbesondere auf Social Media – verbreiteten Hass- und Hetzreden sowie Aufrufen zur Gewalt, wo mit einem Mausklick Massen mobilisiert werden können (Cadwalladr, 2024).

Eine achte globale Gefahr bilden die Risiken für die psychische und physische Gesundheit von Menschen (Mahari & Pataranutaporn, 2024), insbesondere von Kindern und Jugendlichen, aufgrund der Auswirkungen von DS, u.a. von Social Media. Unter anderem prüft die WHO, Social Media-Nutzung als Sucht anzuerkennen (Montag, 2023), weil Social Media folgende Züge besitzt: «The internet is a giant hypodermic, and the content, including social media like Meta, are the psychoactive drugs» (Greenfield, 2023).

Eine neunte globale Gefahr sind Cyberangriffe (Kirchschräger, 2021). Wachsende Digitalisierung und eine Steigerung des Einsatzes von DS im geschäftlichen und privaten Leben von Menschen

sowie von systemischer Infrastruktur von Gesellschaften bringen eine Zunahme der Verletzbarkeit für Cyberangriffe mit sich. Cyberkriminalität und ihre Bekämpfung belasten ökonomisch Volkswirtschaften massiv.

Eine zehnte globale Gefahr bilden die militärischen Anwendungen von DS. Automatisierte Waffensysteme sind nicht in der Lage, in bewaffneten Konflikten zwischen Kombattant:innen und Nicht-Kombattant:innen zu unterscheiden und das Prinzip der Verhältnismässigkeit (Asaro, 2008) anzuwenden, was zu mehr Fehlhandlungen und Verbrechen führt (Sharkey, 2011). Zudem spricht auch gegen automatisierte Waffen, dass sie zu mehr Kriegen führen würden (Kahn, 2017), weil u.a. eine geringere Zahl von Opfern zu erwarten wäre (Wallach, 2015) und somit ein geringerer politischer Preis gezahlt werden müsste (Wagner, 2014).

## Menschenrechtsbasierte DS und Internationale Agentur für DS (IDA)

### *Konkreter Handlungsbedarf*

Es ist notwendig, ethische Chancen und ethische Risiken frühzeitig und präzise zu identifizieren, um die ethischen Vorteile von DS nachhaltig nutzen zu können und die ethischen Nachteile von DS zu bewältigen bzw. zu vermeiden. Die Menschen müssen aktiv werden, damit Digitalisierung und DS nicht einfach geschehen, sondern die Menschen sie gestalten. Dies ist notwendig, damit Digitalisierung und DS nicht auf ein Instrument reduziert werden, das der reinen Effizienz dient, sondern ihr ethisches Potenzial entfalten können. Die wirtschaftlichen Eigeninteressen von einigen wenigen multinationalen Technologiekonzernen, die den digitalen Wandel und DS bisher fast ausschliesslich steuern, sind von einem ethischen Standpunkt als alleiniges Ziel im digitalen Bereich und im Bereich DS zu überwinden. Aus ethischer Sicht gilt es, den Menschen und dem Planeten ein nachhaltiges Gedeihen zu ermöglichen und zu gewährleisten, dass die Menschenrechte weltweit nicht nur offline, sondern auch online und in der digitalen Sphäre sowie im Bereich von DS eingehalten und umgesetzt werden. Elon Musk (Mitgründer von Paypal, Chef von Tesla

und SpaceX) – eine Stimme, die aus ethischer Perspektive für vieles kritisiert werden kann, aber sich sicherlich nicht den Vorwurf gefallen lassen muss, technologiefeindlich zu sein – warnt: “AI is far more dangerous than nukes [nuclear warheads]. Far. So why do we have no regulatory oversight? This is insane.” (Clifford, 2018)

Der Astrophysiker Stephen Hawking, der nur dank Technologien mit der Menschheit kommunizieren konnte, hebt hervor: “Unless we learn how to prepare for, and avoid, the potential risks, AI could be the worst event in the history of our civilization. It brings dangers, like powerful autonomous weapons, or new ways for the few to oppress the many. It could bring great disruption to our economy.” (Kharpal, 2017) Es besteht also konkreter Handlungsbedarf. Im Folgenden sollen zwei konkrete Handlungsvorschläge skizziert werden, die beide auf der globalen Ebene angesiedelt sind, da es sich bei Digitalisierung und bei DS um ein globales Phänomen handelt. Die Priorität wird auf die Regulierung von DS und die Durchsetzung der Regulierung von DS gesetzt, um dem Ausmass der ethischen Chancen und ethischen Gefahren von DS gerecht zu werden. Um einem Missverständnis vorzubeugen: Damit soll nicht gesagt werden, dass

nicht andere Vorgehensweisen wie beispielsweise «ethics by design», die auf der individuellen Ebene der Softwareingenieur:innen und -entwickler:innen ansetzen und diese zur Verantwortung ziehen, ebenfalls wichtig sind. Natürlich ist es sinnvoll, wenn bereits bei der Entwicklung und dem Design von DS ethische Vorgaben beachtet werden. Angesichts der enormen wirtschaftlichen Interessen und damit verbundenen Marktzwänge, welche die einzelnen Softwareingenieur:innen und -entwickler:innen entsprechend unter Druck setzen, wird dies allein aber nicht ausreichen und würde die betroffenen Individuen überfordern. Um es mit einer Analogie zu veranschaulichen: Nukleartechnologien überlässt man auch nicht nur «ethics by design» ...

### ***Menschenrechtsbasierte DS***

Einen ersten konkreten Handlungsvorschlag stellen menschenrechtsbasierte datenbasierte Systeme HRBDS (Kirchschräger, 2021) dar. Ausgehend von den bisherigen Überlegungen sollte eine menschenrechtsbasierte Gestaltung, Entwicklung, Produktion, Nutzung und Nichtnutzung datenbasierter Systeme angestrebt werden. HRBDS bedeutet mit anderen Worten, dass die Menschenrechte im gesamten Lebenszyklus von DS und in der gesamten



Wertschröpfungskette von DS geachtet, geschützt, umgesetzt und verwirklicht werden (d.h. bei der Schürfung der für DS notwendigen Rohstoffe, der Gestaltung, der Entwicklung, der Herstellung, dem Vertrieb, der Nutzung oder der Nichtnutzung von DS aufgrund von Menschenrechtsbedenken).

Um es an einem konkreten Beispiel zu verdeutlichen, bedeutet HRBDS, dass z. B. die Menschenrechte auf Privatsphäre und Datenschutz in ihrer Relevanz für die Menschenwürde und Freiheit des Menschen für alle Menschen verteidigt werden müssen – was auch ausschliesst, dass Menschen sich und ihre Daten sowie ihre Privatsphäre als Produkte verkaufen können sollen. Aktuelle Geschäftsmodelle von Videokonferenzsoftware wie z. B. ZOOM (Laaff, 2020) überwachen die Nutzer:innen und verletzen ihr Menschenrecht auf Privatsphäre und Datenschutz, indem sie ständig ihre Daten sammeln, generieren und an Dritte verkaufen, obwohl unsere Privatsphäre und unsere Daten nach dem Prinzip der Unveräußerlichkeit nicht zum Verkauf stehen sollten. Mit anderen Worten, es muss möglich sein, mit der Bereitstellung und Förderung einer Videokonferenzsoftware ein profitables Geschäftsmodell zu schaffen, das keine Menschenrechtsverletzungen impliziert, indem es

keine Daten seiner Nutzer:innen sammelt und generiert und diese nicht an Dritte verkauft.

HRBDS eröffnet mit «zweckgebundener Datenverwendung» die Möglichkeit, Daten intensiv zu nützen und gleichzeitig die Menschenrechte zu achten. Beim automatisierten Fahren zum Beispiel muss sich der Mensch mit seinen Daten identifizieren und seinen Standort und sein Ziel eingeben, um überhaupt in den Genuss des automatisierten Fahrens zu kommen. Diese Daten werden aber nur bereitgestellt, um den Fahrvorgang zu ermöglichen. Sie dürfen weder für andere Zwecke verwendet noch an Dritte weiterverkauft werden. Die Nutzer:innen haben auch nicht die Möglichkeit, diese Daten selbst zu verkaufen (z. B. um einen Rabatt zu erhalten). Darüber hinaus dürfen vollständig anonymisierte Daten nur mit der informierten Zustimmung der Nutzer:innen zur Optimierung der gemeinsamen automatisierten Mobilität aufbewahrt werden. Eine Analogie zeigt: Es funktioniert! Wenn man zur Ärztin geht, teilt man auch persönliche Daten mit, damit die Ärztin weiss, wen sie vor sich hat, und man erzählt ihr von seiner Krankheit, um hoffentlich Leidenslinderung und Heilung zu erfahren, ohne dass entweder die Ärztin diese Daten weiterverkaufen darf oder den

Patient:innen angeboten wird, diese Daten zu verkaufen, um eine bessere medizinische Behandlung zu erhalten. Die Ärztin kann auch die Patient:innenakte mit der Krankengeschichte streng vertraulich behandeln – ausschliesslich zum Zwecke einer besseren Behandlung der Patient:innen. Es ist auch möglich, vollständig anonymisierte Daten zu Forschungszwecken weiterzugeben, wenn die Patient:innen in Kenntnis der Sachlage ihre Zustimmung dazu geben.

### ***Internationale Agentur für datenbasierte Systeme (IDA)***

Ein zweiter konkreter Handlungsvorschlag, der auch die praktische Umsetzung von HRBDS fördern soll, umfasst die Schaffung einer Internationalen Agentur für datenbasierte Systeme (International Data-Based Systems Agency IDA) – analog zur Internationalen Atomenergiebehörde (IAEA) ([www.iaea.org](http://www.iaea.org)) – bei der UNO. IDA soll als Plattform für die technische Zusammenarbeit im Bereich der Digitalisierung und DS für staatliche und nicht-staatliche Akteure dienen (natürlich einschliesslich des privaten Sektors und der Zivilgesellschaft sowie Organisationen und Institutionen, die [bereits] in diesem Bereich tätig sind, u.a. ITU, die Broadband Commission, UNESCO, UN DESA),

sowie die Menschenrechte, Sicherheit und friedliche Nutzung von DS fördern. Zudem soll sie eine globale Aufsichts- und Überwachungsinstitution und Regulierungsbehörde im Bereich der digitalen Transformation und DS sein, die für den Prüfungsprozess zur Marktzulassung zuständig ist. Eingebunden in die UNO sollte sie sich für die sichere und friedliche Nutzung von DS einsetzen und so zu internationalem Frieden und Sicherheit, zur Achtung und Verwirklichung der Menschenrechte und zu den Zielen für nachhaltige Entwicklung der UNO beitragen. Ihr globaler und umfassender Ansatz soll es ihr ermöglichen, die Gefahr der Fragmentierung in diesem Bereich zu bewältigen.

Warum IDA? Lässt sich mit den ethischen Chancen und Risiken von Digitalisierung und DS nicht anders umgehen? Bestehen in diesem Bereich nicht bereits genug Initiativen? In der Tat haben zahlreiche Erklärungen, Empfehlungen, Kataloge von Prinzipien und Richtlinien die Debatte über die internationale Governance von DS bereichert. Sie können als erste Generation von Governance-Initiativen bezeichnet werden: «the sermons» (Kirchschräger, 2024c). Dazu zählen zahlreiche Vorhaben von Staaten und Zivilgesellschaft auf nationaler,

regionaler und internationaler Ebene sowie Initiativen, die aus berufsethischen Kontexten hervorgehen (ebd.). Die grösste Herausforderung bei diesen Erklärungen, Empfehlungen, Grundsätzen und Leitlinien ist der Schritt von der Theorie zur Praxis. Sie bleiben maximal «soft law» “which is a tool often used to either avoid or anticipate formal legislation” (Nevejans, 2016, S. 26).

Darüber hinaus hat die EU den European Union’s Artificial Intelligence Act (‘EU AI Act’) (European Parliament, 2023) verabschiedet. Der EU AI Act soll einen umfassenden Rechtsrahmen für die Regulierung von DS in der gesamten EU darstellen, die Sicherheit und die Achtung der Grundrechte durch DS gewährleisten sowie Investitionen und Innovationen im Bereich der DS fördern.

Andere rechtliche Initiativen, die zusammen mit dem EU AI Act die zweite Generation von Governance-Initiativen bilden – «the locals» – (Kirchschräger, 2024c), werden auch in China und den USA verfolgt, zu denen auch neue Vorschriften zählen, die mehrere Regierungen auf bundesstaatlicher Ebene in den USA erlassen haben, fliessen in die Kategorisierung als “American Market-Driven

Regulatory Model”, “Chinese State-Driven Regulatory Model”, and “European Rights-Driven Regulatory Model” (Bradford, 2023) ein.

Schliesslich kennt die Debatte über Global Governance in DS auch die dritte Generation von Governance-Initiativen: «the international players» (Kirchschräger, 2024c). Dazu gehören die folgenden zwei Hauptvorschläge: das Modell des Intergovernmental Panel on Climate Change (IPCC) (Carnegie Council for Ethics in International Affairs, 2023) und das Modell der International Civil Aviation Organization (ICAO) (Baker & McKenzie International, 2023).

Das Modell des Intergovernmental Panel on Climate Change (IPCC) besteht aus einem Expertengremium. Der IPCC wurde 1988 von der UNO gegründet und umfasst Mitgliedsländer aus der ganzen Welt. Er versorgt die Regierungen mit wissenschaftlichen Informationen, die sie für die Entwicklung von Klimapolitiken nutzen können. Das Expertengremium auf dem Gebiet der DS würde politischen Entscheidungsträger:innen und Regierungen Informationen, Szenarien und Modelle für ihre Entscheidungsfindung liefern.

Das Modell der International Civil Aviation Organization (ICAO) besteht aus einem verbindlichen globalen Rahmen und dessen Umsetzung. Die ICAO ist eine Einrichtung der UNO und hat ihre Grundlage im Abkommen über die internationale Zivilluftfahrt. Die ICAO ist das globale Forum der Staaten für die internationale Zivilluftfahrt. Sie entwickelt Richtlinien und Standards, bietet Audits zur Einhaltung der Vorschriften, Studien, Analysen und Unterstützung für Staaten und Interessengruppen an und trägt zur weltweiten Angleichung der Luftverkehrsvorschriften bei.

Während es «the sermons» an der Umsetzung unter der Arbeitswoche mangelt und «the locals» einen nationalen oder regionalen Fokus haben, obwohl DS ein globales Phänomen darstellen, verorten «the international players» die Governance von DS auf internationaler Ebene und gehen sie auf angemessene Weise institutionell an, um in der Realität und vor Ort eine positive Wirkung zu erzielen.

Gleichzeitig ist zu bedenken, dass das IPCC-Modell nicht die gewünschte Wirkung vor Ort erzielen

wird, da es weder über rechtliche Befugnisse noch über rechtliche Durchsetzungsinstrumente verfügt. Diese Schwäche des IPCC-Modells wird deutlich, wenn man dem UN-Hochkommissar für Menschenrechte, Volker Türk (2023), zuhört: “Victims and experts (...) have raised the alarm bell for quite some time, but policy makers and developers of AI have not acted enough – or fast enough – on those concerns. We need urgent action by governments and by companies. And at the international level, the United Nations can play a central role in convening key stakeholders and advising on progress. There is absolutely no time to waste. The world waited too long on climate change. We cannot afford to repeat that same mistake”.

Hinsichtlich des ICAO-Modells ergeben sich Zweifel, ob man eine in sich geschlossene Branche wie die Luftfahrt mit einer Querschnittstechnologie wie DS vergleichen kann. Letztere wirft vielfältige und vielschichtige rechtliche und ethische Fragen auf (Baker & McKenzie International, 2023).

Angesichts der ethischen Chancen und Gefahren von DS werden die Grenzen der bisherigen drei



Generationen von Governance-Initiativen deutlich, wenn man sich dieser u.a. auch riskanten Realität mit den folgenden Testfragen für die bereits existierenden «the sermons», «the locals» und «the international players» nähert:

- Können menschenrechtsverletzende staatliche Handlungen oder Geschäftspraktiken aufgrund dieser Vorschläge konkret unterbunden werden?
- Müssen multinationale Technologieunternehmen aufgrund dieser Vorschläge etwas an ihren menschenrechtsverletzenden Geschäftspraktiken ändern?
- Können Staaten oder multinationale Technologieunternehmen aufgrund dieser Vorschläge für ihre Menschenrechtsverletzungen zur Rechenschaft gezogen werden?

Die bereits vorhandenen Generationen von Governance-Initiativen erhalten in allen drei Fällen negative Antworten. Daher ist es notwendig, darüber nachzudenken, wie die Stärken von «the sermons», «the locals» und «the international players» kombiniert und ihre Schwächen vermieden werden können, um die globalen Bedrohungen von DS auf wirksame und globale Weise anzugehen und

gleichzeitig von den ethischen Möglichkeiten von DS zu profitieren. Die IDA soll dies erreichen.

Die IDA soll nach dem Vorbild der Internationalen Atomenergiebehörde (IAEA) als «Institution mit Zähnen» aufgebaut werden, da die IAEA dank ihrer rechtlichen Befugnisse, Funktionen, Durchsetzungsmechanismen und Instrumente in der Lage war, Innovationen und ethische Chancen zu fördern und gleichzeitig die Menschheit und den Planeten vor den existenziellen Risiken im Bereich der Nukleartechnologien zu schützen, die denselben doppelten Charakter wie DS haben, d.h. sowohl ethische Vor- als auch Nachteile aufweisen. Die Erfahrungen der IAEA unterstreichen, wie wichtig es ist, den Menschenrechten, der Transparenz und der Rechenschaftspflicht im Rahmen der Governance von DS Priorität einzuräumen. Durch die Kombination eines menschenrechtsbasierten Ansatzes für die Entwicklung und den Einsatz von DS mit der Einrichtung der IDA als zentraler Regulierungsbehörde kann die internationale Gemeinschaft existenzielle Risiken von DS proaktiv angehen und gleichzeitig das transformative Potenzial von DS zur Förderung von Menschenrechten und Nachhaltigkeit nutzen. Auf

der Grundlage der Lehren, die wir aus den Nukleartechnologien und der Gründung der IAEA gezogen haben, stellt die Einrichtung der IDA einen gangbaren Weg zu einer wirksamen globalen Kontrolle existenzieller DS-Risiken dar und gewährleistet eine verantwortungsvolle und ethische Entwicklung von DS zum Wohle der Menschheit und des Planeten.

Die Realisierung von IDA ist realistisch, weil die Menschheit in der Vergangenheit gezeigt hat, dass sie sich, wenn das Wohlergehen der Menschen und des Planeten auf dem Spiel steht, auf das technisch Gesollte konzentrieren kann, anstatt blindlings alles technisch Machbare zu verfolgen.

Die Menschheit hat die Nukleartechnologie verfolgt, die Atombombe entwickelt und sie sogar mehr als einmal eingesetzt. Doch um noch Schlimmeres zu verhindern, hat die Menschheit dann die Forschung und Entwicklung der Kerntechnik trotz des überwältigenden Widerstands staatlicher und nichtstaatlicher Akteure massiv eingeschränkt. Dass nichts Schlimmeres passiert ist, ist vor allem internationalen Richtlinien, konkreten Durchsetzungsmechanismen und der Internationalen Atomenergiebehörde (IAEA) der UNO zu verdanken.

Auch bei den Fluorchlorkohlenwasserstoffen (FCKW) hat sich die Menschheit im Rahmen des Montrealer Protokolls von 1987 (The Montreal Protocol, 1987) dazu entschlossen, ozonschichtschädigende Stoffe zu verbieten und das Verbot konsequent durchzusetzen. Auch hier war der Widerstand gross, u.a. aufgrund von Eigeninteressen der Privatwirtschaft. Diese Regelung und ihre kompromisslose Durchsetzung haben dazu geführt, dass sich das Ozonloch nun langsam schliesst.

### ***Globales positives Echo für HRBDS und IDA***

Zudem stossen HRBDS und IDA weltweit auf positiven Anklang. Neben einem wachsenden internationalen und interdisziplinären Netzwerk von Expert:innen, das menschenrechtsbasierte DS und die Einrichtung von IDA fordert, haben «the Elders» – eine von Nelson Mandela gegründete unabhängige Gruppe führender Persönlichkeiten der Welt, der auch der ehemalige UN-Generalsekretär Ban Ki-moon, Ellen Johnson Sirleaf (ehemalige Präsidentin von Liberia und Afrikas erstes gewähltes weibliches Staatsoberhaupt sowie Friedensnobelpreisträgerin) und die erste weibliche Präsidentin Irlands, Mary Robinson, angehören –, die konkreten Empfehlungen für menschenrechtsbasierte DS und die Errichtung einer globaler

Institution bei der UNO aufgegriffen und die UNO aufgefordert, entsprechende Massnahmen zu ergreifen. So verlangen «the Elders» (2023): «A new global architecture is needed to manage these powerful technologies within robust safety protocols, drawing on the model of the Nuclear Non-Proliferation Treaty and the International Atomic Energy Agency. These guardrails must ensure AI is used in ways consistent with international law and human rights treaties. AI's benefits must also be shared with poorer countries. No existing international agency has the mandate and expertise to do all this. The Elders now encourage a country or group of countries to request as a matter of priority, via the UN General Assembly, that the International Law Commission draft an international treaty establishing a new international AI safety agency».

UNO-Generalsekretär António Guterres befürwortet auch die Schaffung einer internationalen KI-Überwachungsstelle nach dem Vorbild der Internationalen Atomenergiebehörde (IAEA): “I would be favorable to the idea that we could have an artificial intelligence agency (...) inspired by what the international agency of atomic energy is today.”

(Nichols, 2023) Er hat im UN-Sicherheitsrat am 18. Juli 2023 ein neues UN-Gremium gefordert, um Bedrohungen durch künstliche Intelligenz zu bekämpfen (Guterres, 2023).

UNO-Hochkommissar für Menschenrechte Volker Türk (2023) forderte in seiner Stellungnahme zu KI und Menschenrechten am 12. Juli 2023 im UNO-Menschenrechtsrat «dringende Massnahmen» und schlug menschenrechtsbasierte DS und eine koordinierte globale Reaktion in Form einer institutionellen Lösung vor.

Der UNO-Menschenrechtsrat (2023) verabschiedete am 14. Juli 2023 einstimmig eine Resolution zum Thema «New and emerging digital technologies and human rights», in der zum ersten Mal ausdrücklich auf KI und die Förderung und den Schutz der Menschenrechte Bezug genommen wird. Die UN-Resolution betont, neue und aufkommende Technologien, die Auswirkungen auf die Menschenrechte haben, «may lack adequate regulation», unterstreicht den «need for effective measures to prevent, mitigate and remedy adverse human rights impacts of such technologies» und

hebt die Notwendigkeit hervor, die Menschenrechte «throughout the lifecycle of artificial intelligence systems» zu achten, zu schützen und zu fördern. Sie fordert einen Rahmen für Folgenabschätzungen im Zusammenhang mit den Menschenrechten, eine Sorgfaltspflicht zur Bewertung, Verhinderung und Abschwächung negativer Auswirkungen auf die Menschenrechte sowie die Gewährleistung wirksamer Abhilfemassnahmen und menschlicher Aufsicht, Rechenschaftspflicht und rechtlicher Verantwortung.

Am 21. März 2024 verabschiedete die UNO-Generalversammlung (2024) einstimmig die Resolution «Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development», die der nachhaltigen Entwicklung für alle zugutekommen sollen. Darin hebt die internationale Staatengemeinschaft hervor: «The same rights that people have offline must also be protected online, including throughout the life cycle of artificial intelligence systems.»

Papst Franziskus (2024) fordert dazu auf, «einen verbindlichen internationalen Vertrag zu schliessen,

der die Entwicklung und den Einsatz von künstlicher Intelligenz in ihren vielfältigen Formen regelt», und unterstützt die zwei konkreten Handlungsvorschläge menschenrechtsbasierte DS und Schaffung einer Internationalen Agentur für datenbasierte Systeme (IDA) bei der UNO zur Durchsetzung der Regulierung für DS.

Auch einige Stimmen aus multinationalen Technologieunternehmen – unter anderem Sam Altman erneut am World Economic Forum WEF in Davos 2024 (Santelli, 2024) – haben die Schaffung von IDA gefordert.

Als Bürger:innen, als Konsument:innen und als Nutzer:innen tragen wir ebenfalls eine Verantwortung, die Digitalisierung und die Entwicklung, Produktion und den Einsatz von DS nicht einfach geschehen zu lassen, sondern mitzugestalten. Natürlich ist unsere Verantwortung im Vergleich zur Verantwortung von Staaten und Unternehmen kleiner korrespondierend mit unserer kleineren Macht und unserem geringeren Einfluss auf diese technologiebasierte Entwicklung, aber sie besteht und existiert. So sollten wir uns politisch für einen



nachhaltigen und menschenwürdigen technologischen Fortschritt einsetzen. Wir sollten uns als «Konsum Actors» (Kirchschräger 2016b) im Rahmen unser Kaufentscheidungen für eine bessere Welt engagieren, denn diese stellen immer auch eine politische Stimmabgabe dar für oder gegen Menschen und Umwelt ausbeutende Lieferketten und Geschäftspraktiken. Und wir könnten menschenrechtsbasierte DS und die Schaffung von IDA bei der UNO ideell unterstützen:

<https://idaonline.ch/join-ida/>

## Literaturverzeichnis

- Asaro, P. M. (2008). How just could a robot war be?. In Adam Briggles, Katinka Waelbers, Philip A. E. Brey, (Hrsg.), *Current Issues in Computing and Philosophy*. IOS Press: Amsterdam. S. 50-64.
- Atz, K., Cotos, L., Isert, C., Håkansson, M., Focht, D., Hilleke, M., Nippa, D. F., Iff, M., Ledergerber, J., Schiebroek, C. C. G., Romeo, V., Hiss, J. A., Merk, D., Schneider, P., Kuhn, B., Grether, U. & Schneider, G. (2024). Prospective de novo drug design with deep interactome learning. *Nature Communications*, 15(3408).  
DOI: 10.1038/s41467-024-47613-w.
- Baker & McKenzie International. (2023, 8. November). Can a global framework regulate AI Ethics? Insight Plus.  
<https://insightplus.bakermckenzie.com/bm/investigations-compliance-ethics/international-can-a-global-framework-regulate-ai-ethics> (Abgerufen am 8.8.2024).
- Boddington, P. (2023). *AI Ethics*. Springer: Cham.

Bradford, A. (2023). *Digital Empires. The Global Battle to Regulate Technology*. Oxford University Press: Oxford, S. 35–145.

Cadwalladr, C. (2024, 3. August). ‘A polarisation engine’: how social media has created a ‘perfect storm’ for UK’s far-right riots. *The Guardian*. <https://www.theguardian.com/media/article/2024/aug/03/a-polarisation-engine-how-social-media-has-created-a-perfect-storm-for-uks-far-right-riots> (Abgerufen am 8.8.2024).

Cambridge Dictionary. (o. D.). Digital. <https://dictionary.cambridge.org/dictionary/english/digital> (Abgerufen am 8.8.2024).

Carnegie Council for Ethics in International Affairs. (2023). *Envisioning Modalities for AI Governance: A Response from AIEI to the UN Tech Envoy*. Artificial Intelligence & Equality Initiative. <https://www.carnegiecouncil.org/media/article/envisioning-modalities-ai-governance-tech-envoy#gai0> (Abgerufen am 8.8.2024).

Clifford, C. (2018, 13. März). Elon Musk: “Mark my words – A.I. is far more dangerous than nukes”. CNBC.

<https://www.cnn.com/2018/03/13/elon-musk-at-sxsw-a-i-is-more-dangerous-than-nuclear-weapons.html> (Abgerufen am 8.8.2024).

Dreyfus, H. L. & Dreyfus, S. E. (1986). *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. Free Press: New York.

Dreyfus, H.L. (1972). *What Computers Can't Do: The Limits of Artificial Intelligence*. MIT Press: New York.

European Parliament. (2023). Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI. <https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai> (Abgerufen am 8.8.2024).

Greenfield, D. (2023, 25. Oktober). Is Social Media Addictive? Here's What the Science Says. New York Times.

<https://www.nytimes.com/2023/10/25/health/social-media-addiction.html> (Abgerufen am 8.8.2024).

Gueterres, A. (2023, 18. Juli). Secretary-General's remarks to the Security Council on Artificial Intelligence. UN Secretary General.

<https://www.un.org/sg/en/content/sg/speeches/2023-07-18/secretary-generals-remarks-the-security-council-artificial-intelligence> (Abgerufen am 8.8.2024).

Heikkilä, M. (2022, 12. Dezember). The viral AI avatar app Lensa undressed me – without my consent. MIT Technology Review.

<https://www.technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-undressed-me-without-my-consent> (Abgerufen am 8.8.2024).

Kahn, L. (2017). "Military Robots and the Likelihood of Armed Combat". In Patrick Lin, Keith Abney, Ryans Jenkins (Hrsg.), Robot Ethics 2.0: From

- Autonomous Cars to Artificial Intelligence. New York: Oxford University Press, S. 274-287.
- Kant, I. (1974). Grundlegung zur Metaphysik der Sitten. In W. Weischedel (Hrsg.), Werkausgabe (7. Aufl.). Suhrkamp: Frankfurt am Main.
- Kharpal, A. (2017, 6. November). Stephen Hawking says A.I. could be ‘worst event in the history of our civilization’. CNBC.  
<https://www.cnbc.com/2017/11/06/stephen-hawking-ai-could-be-worst-event-in-civilization.html> (Abgerufen am 8.8.2024).
- Kirchschräger, P. G. (2013a). *Wie können Menschenrechte begründet werden?* ReligionsRecht im Dialog (Bd. 15). LIT Verlag: Berlin.
- Kirchschräger, P.G. (2013b). Human Rights as an Ethical Basis for Science. *Journal of Law, Information and Science*, 22(2), S. 1–17.
- Kirchschräger, P. G. (2016a). *Menschenrechte und Religionen*. Gesellschaft-Ethik-Religion (Bd. 7). Ferdinand Schöningh: Paderborn.

Kirchschräger, P. G. (2016b). KonsumActors – mehr Macht beim Einkauf als an der Urne? Konsumethische Überlegungen zur Verantwortung beim Einkaufen. *ETHICA*, 24(2), S. 133–157.

Kirchschräger, P. G. (2020). Human Dignity and Human Rights: Fostering and Protecting Pluralism and Particularity. *Interdisciplinary Journal for Religion and Transformation in Contemporary Society*, S. 90–106.

Kirchschräger, P. G. (2021). *Digital Transformation and Ethics. Ethical Considerations on the Robotization and Automation of Society and the Economy and the Use of Artificial Intelligence*. Nomos: Baden-Baden.

Kirchschräger, P. G. (2022). Ethische KI? Datenbasierte Systeme (DS) mit Ethik. *HMD Praxis der Wirtschaftsinformatik*, 59(2), S. 482–494.

Kirchschräger, P. G. (2023). *Ethisches Entscheiden*. Nomos: Baden-Baden.

Kirchschräger, P. G. (2024a). In an Era of Digital Disruptions, Ethics Can't Be an Afterthought – Part

1 and 2. Business and Human Rights Journal Blog:  
Cambridge University Press.

<https://www.cambridge.org/core/blog/2024/04/11/in-an-era-of-digital-disruptions-ethics-cant-be-an-afterthought/> (Abgerufen am 8.8.2024).

Kirchschräger, P. G. (2024b). Für eine menschenwürdige und nachhaltige Zukunft. «Datenbasierte Systeme (DS)» statt «Künstliche Intelligenz (KI)». In Claudia Hamm (Hrsg.), *Automatensprache*. Hanser Akzente, 71(3/2023 & 1/2024), 156-168. Hanser Verlag: München.

Kirchschräger, P. G. (2024c). An International Data-Based Systems Agency IDA: Striving for a Peaceful, Sustainable, and Human Rights-Based Future. *Philosophies*, 9(3), S. 73. DOI:

<https://doi.org/10.3390/philosophies9030073>.

Laaff, M. (2020, 31. März). Ok, Zoomer, *Zeit Online*.

[https://www.zeit.de/digital/2020-](https://www.zeit.de/digital/2020-03/videokonferenzen-zoom-app-homeoffice-quarantaene-coronavirus)

[03/videokonferenzen-zoom-app-homeoffice-](https://www.zeit.de/digital/2020-03/videokonferenzen-zoom-app-homeoffice-quarantaene-coronavirus)

[quarantaene-coronavirus](https://www.zeit.de/digital/2020-03/videokonferenzen-zoom-app-homeoffice-quarantaene-coronavirus) (Abgerufen am 8.8.2024).



Lensa AI. (o. D.) Lensa App. <https://lensa.app/>  
(Abgerufen am 8.8.2024).

Mahari, R. & Pataranutaporn, P. (2024, 5. August). We need to prepare for ‘addictive intelligence’. MIT Technology Review.  
<https://www.technologyreview.com/2024/08/05/1095600/we-need-to-prepare-for-addictive-intelligence/> (Abgerufen am 8.8.2024).

Montag, C. (2023, 19. Oktober). Was verändern Smart Phones in unserem Denkapparat? [Vorlesung]. In ISE-Ringvorlesung «Ethik der digitalen Transformation. Aktuelle Fragen», Universität Luzern.

Nevejans, N. (2016). European Civil Law Rules in Robotics. Study for the JURI Committee. European Parliament: Strasbourg, S. 26.  
[https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL\\_STU\(2016\)571379\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf) (Abgerufen am 8.8.2024).

Nichols, M. (2023, 12. Juni). UN Chief Backs Idea of Global AI Watchdog Like Nuclear Agency.

Reuters.

<https://www.reuters.com/technology/un-chief-backs-idea-global-ai-watchdog-like-nuclear-agency-2023-06-12/> (Abgerufen am 8.8.2024).

Nosthoff, A.-V. & Maschewski, F. (2019). *Die Gesellschaft der Wearables. Digitale Verführung und soziale Kontrolle*. Berlin: Nicolai Publishing & Intelligence GmbH.

Ohly, L. (2019). *Ethik der Robotik und der Künstlichen Intelligenz*. Theologisch-Philosophische Beiträge zu Gegenwartsfragen 22. Berlin: Peter Lang.

Papst Franziskus. (2024). “Artificial Intelligence and Peace”. Message of Pope Francis for the 57th World Day of Peace. 1 January 2024.

<https://www.vatican.va/content/francesco/en/messages/peace/documents/20231208-messaggio-57giornatamondiale-pace2024.html> (Abgerufen am 8.8.2024).

Risse, M. (2023). *Political Theory of the Digital Age: Where Artificial Intelligence Might Take Us*. Cambridge University Press.

Santelli, F. (2024). Sam Altman: “In pochi anni l’IA sarà inarrestabile, serve un’agenzia come per l’energia atomica”. La Repubblica.

[https://www.repubblica.it/economia/2024/01/18/news/sam\\_altman\\_in\\_pochi\\_anni\\_lia\\_sara\\_inarrestabile\\_serve\\_unagenzia\\_come\\_per\\_lenergia\\_atmica-421905376/amp/](https://www.repubblica.it/economia/2024/01/18/news/sam_altman_in_pochi_anni_lia_sara_inarrestabile_serve_unagenzia_come_per_lenergia_atmica-421905376/amp/) [4.7.2024].

Sharkey, Noel E. (2011). The automation and proliferation of military drones and the protection of civilians. *Law Innovation Technology*, 3, S. 229-240.

Snow, O. (2022, 7. Dezember). ‘Magic Avatar’ App Lensa Generated Nudes From My Childhood Photos. *Wired*.

<https://www.wired.com/story/lensa-artificial-intelligence-csem/?bxid=5cc9e15efc942d13eb203f10>  
(Abgerufen am 8.8.2024).

The Elders. (2023). The Elders urge global co-operation to manage risks and share benefits of AI.

<https://theelders.org/news/elders-urge-global-co-operation-manage-risks-and-share-benefits-ai>  
(Abgerufen am 8.8.2024).

The Montreal Protocol (1987). About Montreal Protocol. UN Environment Programme (UNEP). <https://www.unep.org/ozonaction/who-we-are/about-montreal-protocol> (Abgerufen am 8.8.2024)

Türk, V. (2023). Artificial intelligence must be grounded in human rights, says High Commissioner. High Level Side Event of the 53rd Session of the Human Rights Council, 12 July 2023. Office of the High Commissioner for Human Rights. <https://www.ohchr.org/en/statements/2023/07/artificial-intelligence-must-be-grounded-human-rights-says-high-commissioner> (Abgerufen am 8.8.2024).

UN-Generalversammlung (2024). Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development. 24 March 2024.

<https://daccess-ods.un.org/access.nsf/Get?OpenAgent&DS=A/78/L.49&Lang=E> (Abgerufen am 8.8.2023).

UNO-Menschenrechtsrat (2023). Resolution New and emerging digital technologies and human rights. No. 41/11.

<https://daccess-ods.un.org/access.nsf/Get?OpenAgent&DS=A/HRC/RES/41/11&Lang=E> (Abgerufen am 8.8.2024).

Wagner, M. (2014). The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems. *Vanderbilt Journal of Transnational Law*, 47, S. 1371-1424.

Wallach, W. (2015). *A Dangerous Master: How to Keep Technology from Slipping Beyond Our Control*. New York: Basic Books.

Završnik, A. & Simončič, K. (Hrsg.) (2023). *Artificial Intelligence, Social Harms and Human Rights*. Palgrave Macmillan: Cham.

Zhang, X., Venkatachalapathy, S., Paysan, D., Schaerer, P., Tripodo, C., Uhler, C. & Shivashankar G. V. (2024). Unsupervised representation learning of chromatin images identifies changes in cell state and tissue organization in DCIS. *Nature Communications*, 15(6112).

DOI: 10.1038/s41467-024-50285-1.

Zuboff, S. (2019). *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power.* London: Public Affairs.



## Impressum

### Zitiervorschlag:

Peter G. Kirchschräger, Digitalisierung und «künstliche Intelligenz» – Maschinen oder Menschen am Steuer?, mit einem Vorwort von Antonius Liedhegener  
(= MA RWP Lectures, Bd. 5) Luzern: Universität Luzern/  
ZRWP 2024 (DOI: 10.5281/zenodo.13777666)

Universität Luzern  
Zentrum für Religion, Wirtschaft und Politik (ZRWP)  
Frohbürgstrasse 3, 6002 Luzern, [www.zrwp.ch](http://www.zrwp.ch)

DOI 10.5281/zenodo.13777666

Lizenz: CC BY 4.0

Redaktion: Silvia Martens