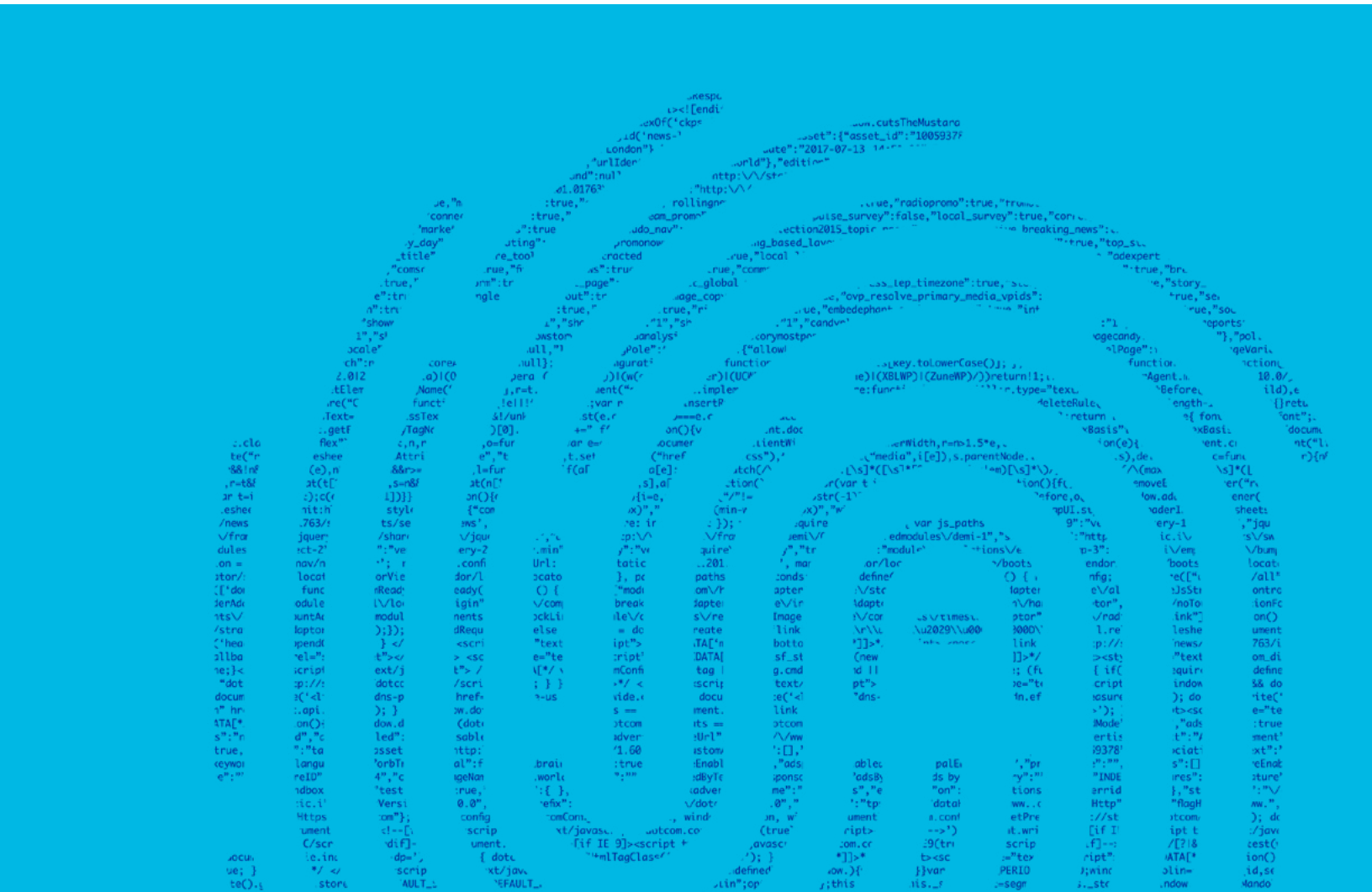
< >

Erklärung von Montréal
Verantwortliche Künstliche Intelligenz_

</ >

ERKLÄRUNG VON MONTREAL FÜR EINE VERANTWORTUNGSVOLLE ENTWICKLUNG VON KÜNSTLICHER INTELLIGENZ

2018



Dieses Dokument ist ein Teil des Berichts zur **ERKLÄRUNG VON MONTREAL FÜR EINE VERANTWORTUNGSVOLLE ENTWICKLUNG VON KÜNSTLICHER INTELLIGENZ** aus dem Jahr 2018. Den vollständigen Bericht finden Sie HIER.

INHALTS-VEZEICHNIS

BEDEUTUNG DER ERKLÄRUNG	5
PRÄAMBEL	7
GRUNDSÄTZE	
1. GRUNDSATZ DES WOHLERGEHENS	8
2. GRUNDSATZ DER ACHTUNG DER AUTONOMIE	9
3. SCHUTZ DES PRIVATLEBENS UND DER PRIVATSPHÄRE	10
4. GRUNDSATZ DER SOLIDARITÄT	11
5. GRUNDSATZ DER DEMOKRATISCHEN PARTIZIPATION	12
6. GRUNDSATZ DER GLEICHHEIT	13
7. GRUNDSATZ DER DIVERSITÄT UND INKLUSION	14
8. GRUNDSATZ DER VORSICHT	15
9. GRUNDSATZ DER VERANTWORTUNG	16
10. GRUNDSATZ DER NACHHALTIGEN ENTWICKLUNG	17
GLOSSAR	18
DANKSAGUNGEN	I
PARTNER	II

WIE DIE ERKLÄRUNG ZU LESEN IST

EINE ERKLÄRUNG – ZU WELCHEM ZWECK?

Die Erklärung von Montreal für eine verantwortungsvolle KI-Entwicklung verfolgt drei Hauptziele:

1. Einen ethischen Rahmen für die Entwicklung und den Einsatz von KI zu erarbeiten
2. Den digitalen Wandel darauf auszurichten, dass alle von der technologischen Revolution profitieren
3. Ein nationales und internationales Diskussionsforum zu eröffnen, um gemeinsam eine gerechte, inklusive und ökologisch nachhaltige KI-Entwicklung zu erreichen

EINE ERKLÄRUNG – WORÜBER?

GRUNDSÄTZE

Das oberste Ziel der Erklärung ist, die ethischen Grundsätze und Werte zu ermitteln, die den grundlegenden Interessen von Menschen und Gruppen dienen. Diese Grundsätze bleiben bei Anwendung auf das Feld des Digitalen und der künstlichen Intelligenz allgemein und abstrakt. Will man sie richtig lesen, ist es wichtig, folgende Punkte zu beachten:

> Auch wenn die Grundsätze in Listenform präsentiert werden, besteht keine Hierarchie unter ihnen. Der zuletzt genannte Grundsatz ist nicht weniger wichtig als der zuerst genannte. Es kann jedoch unter Umständen einem Grundsatz mehr Gewicht beigemessen werden als einem anderen oder ein Grundsatz als relevanter angesehen werden als ein anderer.

> Auch wenn die Grundsätze unterschiedlich sind, müssen sie einheitlich ausgelegt werden, um Widersprüche zu vermeiden, die ihre Anwendung verhindern könnten. Generell gilt die Regel, dass der Anwendungsbereich des einen Grundsatzes begrenzt wird durch den Anwendungsbereich eines anderen Grundsatzes.

> Auch wenn sich in den Grundsätzen die moralische und politische Kultur der Gesellschaft widerspiegelt, in der sie entwickelt wurden, bilden sie die Grundlage für einen interkulturellen und internationalen Dialog.

> Auch wenn sie auf unterschiedliche Art und Weise ausgelegt werden können, ist ihre Auslegung nicht beliebig. Die Grundsätze müssen in jedem Fall kohärent ausgelegt werden.

> Auch wenn es sich um ethische Grundsätze handelt, lassen sich diese in die Sprache der Politik übersetzen und juristisch auslegen.

Aufbauend auf diesen Grundsätze wurden einige Empfehlungen erarbeitet, die als Leitlinien gedacht sind, wie der digitale Wandel innerhalb des ethischen Rahmens der Erklärung verwirklicht werden kann. Es wird auf verschiedene sektorübergreifende Schlüsselthemen eingegangen und überlegt, wie der Übergang zu einer Gesellschaft erfolgen kann, in der KI dazu beiträgt, das Gemeinwohl zu fördern: Algorithmic Governance, digitale Kompetenz, digitale Integration der Diversität und ökologische Nachhaltigkeit.

EINE ERKLÄRUNG – FÜR WEN?

Die Erklärung von Montreal richtet sich an alle Personen, Organisationen und Unternehmen, die an der verantwortungsvollen Entwicklung von künstlicher Intelligenz mitwirken wollen, sei es, um einen wissenschaftlichen oder technologischen Beitrag zu leisten, um gesellschaftliche Projekte zu entwickeln, maßgebliche Regeln (Vorschriften, Kodizes) auszuarbeiten oder auch um falsche (und nicht „schlechte") oder unüberlegte Ansätze erkennen und widerlegen zu können oder in der Lage zu sein, wo notwendig die öffentliche Meinung zu sensibilisieren.

Sie richtet sich auch an die – gewählten oder bestellten – politischen Vertreter; denn die Bürger erwarten von ihnen, dass sie eine Bestandsaufnahme der laufenden gesellschaftlichen Veränderungen vornehmen, dass sie schnell einen Rahmen festlegen, der einen dem Gemeinwohl dienenden digitalen Wandel ermöglicht, und dass sie voraussehen, welche ernst zu nehmenden Risiken die Entwicklung von KI mit sich bringen kann.

EINE ERKLÄRUNG – NACH WELCHER METHODE?

Die Erklärung ist aus einem inklusiven Beratungsprozess entstanden, der den Anstoß gibt zum Dialog zwischen Bürgern, Experten, politischen Entscheidungsträgern, Wirtschaftsvertretern, zivilgesellschaftlichen Organisationen und Berufsverbänden. Dieser Ansatz bietet Vorteile in dreierlei Hinsicht. Er ermöglicht:

1. Soziale und ethische KI-Kontroversen kollektiv zu schlichten
2. Die Qualität der Überlegungen zu verantwortungsvoller KI zu verbessern
3. Die Legitimität der Vorschläge für eine verantwortungsvolle KI zu stärken

Die Ausarbeitung der Grundsätze und Empfehlungen ist eine Gemeinschaftsarbeit, an der zahlreiche Teilnehmer in verschiedensten Foren – im öffentlichen Raum, in Konferenzsälen von Berufsorganisationen, in internationalen Expertenrunden, Forschungsbüros, Unterrichtsräumen oder auch online – mit derselben Strenge mitgewirkt haben.

NACH DER ERKLÄRUNG?

Da sich die Erklärung auf eine sich seit den 1950er Jahren stetig weiterentwickelnde Technologie bezieht, die in exponentiellem Rhythmus bedeutende Innovationen hervorbringt, ist es wichtig, sie als Leitdokument zu betrachten, das für Überarbeitungen und Anpassungen offen ist, um der Entwicklung der KI-Kenntnisse und Techniken sowie dem Nutzerfeedback zur KI-Anwendung in der Gesellschaft Rechnung tragen zu können. Der Abschluss des Ausarbeitungsprozesses für die Erklärung ist gleichzeitig der Startpunkt für eine offene und inklusive Debatte zur Zukunft der sich auf KI-Technologien stützenden Menschheit.

PRÄAMBEL

Erstmals in der Geschichte der Menschheit ist es möglich, autonome Systeme zu schaffen, die komplexe Aufgaben ausführen können, von denen man bislang annahm, dass nur natürliche Intelligenz dazu fähig wäre: Verarbeiten großer Mengen an Daten, Berechnen und Voraussagen, Lernen und Anpassen der der Reaktionen auf sich ändernde Situationen sich ändernde Situationen und Erkennen und Klassifizieren von Objekten. Wegen der immateriellen Natur dieser Aufgaben und in Analogie zur menschlichen Intelligenz bezeichnen wir diese weitreichenden Systeme mit dem allgemeinen Begriff *künstliche Intelligenz*. Künstliche Intelligenz bildet einen wesentlichen Teil des wissenschaftlichen und technologischem Fortschritts und kann erhebliche gesellschaftliche Vorteile mit sich bringen, wie Verbesserung der Lebensbedingungen und Gesundheit, mehr Gerechtigkeit, Schaffung von Wohlstand, Stärkung der öffentlichen Sicherheit und Milderung der Auswirkungen der Aktivitäten des Menschen auf die Umwelt und das Klima. Intelligente Maschinen können nicht nur besser rechnen als Menschen, sie können auch mit fühlenden Wesen interagieren, ihnen Gesellschaft leisten und Betreuungsaufgaben übernehmen.

Die Entwicklung von künstlicher Intelligenz ist jedoch auch mit großen ethischen Herausforderungen und gesellschaftlichen Risiken verbunden. So können intelligente Maschinen die Wahlmöglichkeiten von Personen und Gruppen einschränken, den Lebensstandard verschlechtern, die Organisation von Arbeit und den Stellenmarkt stören, die Politik beeinflussen, mit Grundrechten kollidieren, soziale und wirtschaftliche Ungleichheiten verschärfen und Ökosysteme, das Klima und die Umwelt beeinträchtigen. Auch wenn der wissenschaftliche Fortschritt und das Leben in einer Gesellschaft immer Risiken bergen, ist es Aufgabe der Bürger, die moralischen und politischen Ziele zu bestimmen, die den Risiken, denen wir in einer unsicheren Welt ausgesetzt sind, einen Sinn geben.

Je niedriger die Risiken der Anwendung von künstlicher Intelligenz, desto größer werden ihre Vorteile sein. Die erste Gefahr bei der Entwicklung von KI ist, dass die Illusion entsteht, wir könnten die Herausforderungen der Zukunft durch Berechnungen meistern. Die Gesellschaft auf Zahlenreihen zu reduzieren und sie mit Algorithmen zu regieren ist ein alter Wunschtraum der Menschheit, der noch immer den menschlichen Ehrgeiz beflügelt. Doch wenn es um menschliche Belange geht, ähnelt das Morgen nur selten dem Heute, und Zahlen können weder bestimmen, was moralisch wertvoll, noch was gesellschaftlich wünschenswert ist.

Die Grundsätze dieser Erklärung sind wie Punkte auf einem moralischen Kompass, die es erleichtern, die Entwicklung von künstlicher Intelligenz in Richtung moralisch und gesellschaftlich wünschenswerter Ziele zu lenken. Außerdem bieten sie einen ethischen Rahmen für die Förderung der international anerkannten Menschenrechte in den Bereichen, die von der Verbreitung von künstlicher Intelligenz betroffen sind. In ihrer Gesamtheit legen die formulierten Grundsätze das Fundament, um das gesellschaftliche Vertrauen in KI-Systeme zu pflegen und zu fördern.

Die Grundsätze der Erklärung stützen sich auf die gemeinsame Überzeugung, dass sich der Mensch als soziales Lebewesen, das zu Emotionen, Gedanken und Gefühlen fähig ist, weiterentwickeln will und danach strebt, sein Potenzial auszuschöpfen und seine emotionalen, moralischen und intellektuellen Fähigkeiten frei zu entfalten. Es obliegt den verschiedenen öffentlichen und privaten Akteuren und politischen Entscheidungsträgern auf lokaler, nationaler und internationaler Ebene, dafür zu sorgen, dass die Entwicklung und der Einsatz von künstlicher Intelligenz mit dem Schutz und der vollen Entfaltung der fundamentalen menschlichen Fähigkeiten und Ziele vereinbar sind. In diesem Sinne müssen die vorgeschlagenen Grundsätze kohärent gleichzeitig die spezifischen gesellschaftlichen, kulturellen, politischen und rechtlichen Kontexte ihrer Anwendung berücksichtigt werden.

1

GRUNDSATZ DES **WOHLERGEHENS**

Die Entwicklung und der Einsatz von KI-Systemen müssen dazu beitragen, das Wohlergehen aller fühlenden Lebewesen zu steigern.

1. KI-Systeme müssen den Menschen helfen, ihre Lebensbedingungen, Gesundheit und Arbeitsbedingungen zu verbessern.

2. KI-Systeme müssen den Menschen ermöglichen, ihren persönlichen Präferenzen nachzugehen, solange sie nicht anderen fühlenden Lebenswesen Schaden zufügen.

3. KI-Systeme müssen es Menschen ermöglichen, ihre geistigen und physischen Fähigkeiten einzubringen.

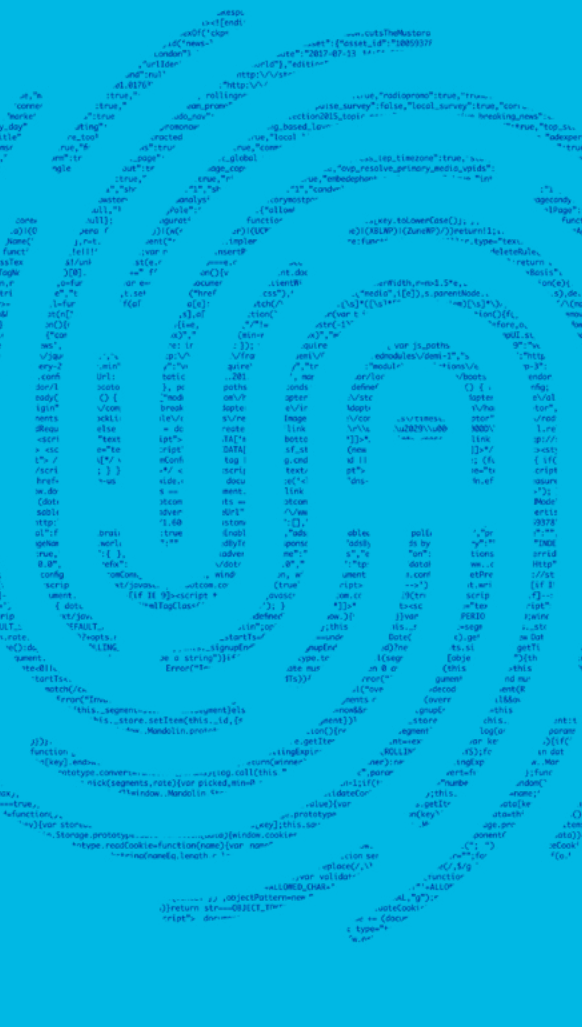
4. KI-Systeme dürfen nicht Grund für Unwohl sein, es sei denn, dadurch können wir ein höheres Wohlergehen erreichen, als dies andernfalls möglich gewesen wäre.

5. Der Einsatz von KI-Systemen darf nicht zu mehr Stress und Angst oder zu einem Gefühl von Belästigung durch die digitale Umgebung führen.

2

GRUNDSATZ DER **ACHTUNG DER AUTONOMIE**

Die Entwicklung und der Einsatz von KI-Systemen müssen unter Beachtung der Selbstbestimmung des Menschen und im Einklang mit dem Ziel, dem Einzelnen mehr Kontrolle über sein Leben und seine Umgebung zu geben, erfolgen.



1. KI-Systeme müssen es dem Einzelnen ermöglichen, seine persönlichen moralischen Ziele und seine Vorstellung von einem lebenswerten Leben zu verfolgen.

2. KI-Systeme dürfen nicht entwickelt oder eingesetzt werden, um den Menschen direkt oder indirekt einen bestimmten Lebensstil aufzuerlegen, indem verpflichtende Überwachungs- und Bewertungs- oder Anreizmechanismen eingeführt werden.

3. Öffentliche Institutionen dürfen KI-Systeme nicht einsetzen, um eine bestimmte Auffassung von dem, was ein gutes Leben ist, zu fördern oder zu diskreditieren.

4. Die Selbstbefähigung der Bürger im Bereich der digitalen Technologien muss gestärkt werden, indem der Zugang zu relevanten Formen des Wissens sichergestellt, das Erlernen fundamentaler Fertigkeiten (digitale Kompetenz und Medienkompetenz) gefördert und die Entwicklung von kritischem Denken unterstützt werden.

5. KI-Systeme dürfen nicht zur Verbreitung von ungesicherten Informationen, Lügen oder Propaganda verwendet werden und müssen so konzipiert sein, dass sie deren Verbreitung verhindern.

6. Bei der Entwicklung von KI-Systemen muss vermieden werden, durch aufmerksamkeitsheischende Techniken oder das Imitieren von menschlichen Charakteristika (Aussehen, Stimme usw.) Abhängigkeiten zu schaffen, die zu Verwechslung zwischen KI-Systemen und Menschen führen könnten.

3

GRUNDSATZ DES **SCHUTZES DES PRIVATLEBENS UND DER PRIVATSPHÄRE**

Privatleben und Privatsphäre müssen vor dem Eindringen von KI-Systemen und von Systemen zur Datenerfassung und Archivierung (DEAS) geschützt werden.

1. Private Räume, in denen Menschen keiner Überwachung oder digitalen Auswertung unterliegen, müssen vor dem Eindringen von KI-Systemen und Systemen zur Datenerfassung und Archivierung (DEAS) geschützt werden.

2. Die Privatheit von Gedanken und Gefühlen muss strikt vor Anwendungen von KI-Systemen und DEAS geschützt werden, die möglicherweise Schaden verursachen könnten, insbesondere im Fall von Anwendungen, die moralische Urteile über Menschen oder deren Entscheidungen in Fragen der Lebensführung fällen.

3. Menschen müssen in ihrem Privatleben immer das Recht auf digitale Abschaltung haben, und KI-Systeme sollten ausdrücklich in regelmäßigen Abständen eine Abschaltoption anbieten und nicht dazu anregen, ununterbrochen verbunden zu bleiben.

4. Menschen müssen weitreichende Kontrollmöglichkeiten über die Informationen zu ihren Präferenzen haben. KI-Systeme dürfen – ohne entsprechende freie und informierte Einwilligung – keine individuellen Präferenzprofile einer Person erstellen, um deren Verhalten zu beeinflussen.

5. DEAS müssen die Vertraulichkeit der Daten und die Anonymität des persönlichen Profils sicherstellen.

6. Jede Person muss in der Lage sein, eine weitreichende Kontrolle über ihre personenbezogenen Daten auszuüben, insbesondere was die Erfassung, Verwendung und Weitergabe der Daten betrifft. Der Zugang einer Person zu KI-Systemen und digitalen Diensten darf nicht davon abhängig gemacht werden, dass sie auf ihre Kontrollmöglichkeiten oder das Eigentum an ihren personenbezogenen Daten verzichtet.

7. Jeder sollte frei sein, seine personenbezogenen Daten an Forschungsorganisationen weiterzugeben, um so zu Fortschritt und Wissensausbau beizutragen.

8. Die Integrität der persönlichen Identität des Einzelnen muss garantiert sein. KI-Systeme dürfen nicht eingesetzt werden, um Aussehen, Stimme oder andere individuelle Merkmale einer Person zu imitieren oder zu verändern mit dem Ziel, Rufschädigung zu betreiben oder andere Personen zu manipulieren.

4

GRUNDSATZ DER **SOLIDARITÄT**

Die Entwicklung von KI-Systemen muss in einer Weise erfolgen, dass die Solidaritätsbande zwischen Menschen und zwischen Generationen aufrechterhalten werden.

1. KI-Systeme dürfen das Bestehen von moralischen und emotionalen menschlichen Beziehungen nicht gefährden, und ihre Entwicklung muss darauf ausgerichtet sein, diese Beziehungen zu stärken und die Verwundbarkeit und Isolation des Einzelnen zu verringern.

2. Bei der Entwicklung von KI-Systemen muss das Ziel sein, bei komplexen Aufgaben mit Menschen zusammenzuarbeiten und die Zusammenarbeit zwischen Menschen zu fördern.

3. KI-Systeme sollten nicht eingeführt werden, um Menschen bei Aufgaben zu ersetzen, die echte menschliche Beziehungen erfordern, sondern Ziel ihrer Entwicklung sollte sein, diese Beziehungen zu fördern.

4. Gesundheitssysteme, die KI-Systeme verwenden, müssen berücksichtigen, dass der Beziehung des Patienten zu seiner Familie und zum Gesundheitspersonal große Bedeutung zukommt.

5. Die Entwicklung von KI-Systemen sollte sollte kein grausames Verhalten gegenüber Robotern stimulieren gegenüber Robotern stimulieren, die Menschen oder Tieren im Aussehen ähneln oder die sich ähnlich wie Menschen oder Tiere verhalten.

6. KI-Systeme sollten dazu beitragen, das Risikomanagement zu verbessern und die Bedingungen für eine Gesellschaft, in der die individuellen und kollektiven Risiken gerechter und ausgeglichener verteilt sind, zu fördern.

5

GRUNDSATZ DER **DEMOKRATISCHEN PARTIZIPATION**

KI-Systeme müssen die Kriterien der Verständlichkeit, Begründbarkeit und Zugänglichkeit erfüllen und einer demokratischen Überprüfung, Debatte und Kontrolle unterzogen werden.

1. KI-Prozesse, die Entscheidungen treffen, die Auswirkungen auf das Leben, die Lebensqualität oder den Ruf einer Person haben, müssen für ihre Entwickler verständlich und nachvollziehbar sein.

2. Von KI-Systemen getroffene Entscheidungen mit Auswirkungen auf das Leben, die Lebensqualität oder den Ruf einer Person sollten immer in einer Sprache nachvollziehbar und begründbar sein, die für die Menschen, die die Systeme anwenden oder den Folgen ihrer Verwendung unterliegen, verständlich ist. Eine Begründung besteht darin, die wichtigsten Faktoren und Parameter für die Entscheidungsfindung transparent zu machen, und sollte dieselbe Form haben wie die Begründung, die wir von einem Menschen erwarten würden, wenn er diese Art von Entscheidung trifft.

3. Der Code der Algorithmen, gleich ob öffentlich oder privat, muss immer für die betreffenden öffentlichen Stellen und Interessengruppen für Überprüfungs- und Kontrollzwecke zugänglich sein.

4. Falls Fehler, unerwartete oder unerwünschte Wirkungen, Sicherheitsverstöße oder Datenlecks in KI-Systemen entdeckt werden, müssen diese zwingend an die zuständigen öffentlichen Stellen, Interessengruppen und betroffenen Personen gemeldet werden.

5. Gemäß den geltenden Transparenzanforderungen für öffentliche Entscheidungen, muss der Code, der von öffentlichen Stellen für Algorithmen zur Entscheidungsfindung verwendet wird, für alle zugänglich sein, mit Ausnahme von Algorithmen, die ein hohes Gefahrenrisiko bei Missbrauch aufweisen.

6. Bei öffentlichen KI-Systemen, die signifikante Auswirkungen auf das Leben der Bürger haben, sollten die Bürger die Möglichkeit und Kompetenz haben, die sozialen Parameter dieser KI-Systeme und deren Ziele und Einsatzgrenzen zu beurteilen und zu entscheiden.

7. Wir müssen jederzeit in der Lage sein zu verifizieren, ob die KI-Systeme tatsächlich das tun, wofür sie programmiert wurden und eingesetzt werden.

8. Eine Person, die einen Dienst nutzt, sollte wissen, ob eine Entscheidung, die sie betrifft oder die Auswirkungen auf sie hat, von einem KI-System getroffen wurde.

9. Wer einen Dienst nutzt, der Chatbots einsetzt, sollte leicht erkennen können, ob er mit einem KI-System oder einer realen Person interagiert.

10. Die Forschung zu künstlicher Intelligenz sollte offen bleiben und allen zugänglich sein.

6

GRUNDSATZ DER **GLEICHHEIT**

Die Entwicklung und der Einsatz von KI-Systemen müssen zur Schaffung einer gerechten Gesellschaft beitragen.

1. KI-Systeme müssen so konzipiert und trainiert werden, dass Diskriminierungen aufgrund sozialer, sexueller, ethnischer, kultureller, religiöser oder sonstiger Unterschiede verhindert und nicht verstärkt oder reproduziert werden.

2. Die Entwicklung von KI-Systemen muss dazu beitragen, Vormachtstellungen unter Gruppen und Menschen, die auf Unterschieden in Macht, Wohlstand oder Wissen basieren, zu beseitigen.

3. Die Entwicklung von KI-Systemen muss allen gesellschaftliche und wirtschaftliche Vorteile bringen und soziale Ungleichheiten und Verwundbarkeiten vermindern.

4. Bei der industriellen Entwicklung von KI-Systemen muss in allen Lebenszyklusphasen, von der Gewinnung der Rohstoffe bis zum Recycling, einschließlich der Datenverarbeitung, auf die Einhaltung von akzeptablen Arbeitsbedingungen geachtet werden.

5. Die digitalen Aktivitäten der Nutzer von KI-Systemen und digitalen Diensten sollten als Arbeit anerkannt werden, die zum ordnungsgemäßen Funktionieren von Algorithmen beiträgt und Mehrwert schafft.

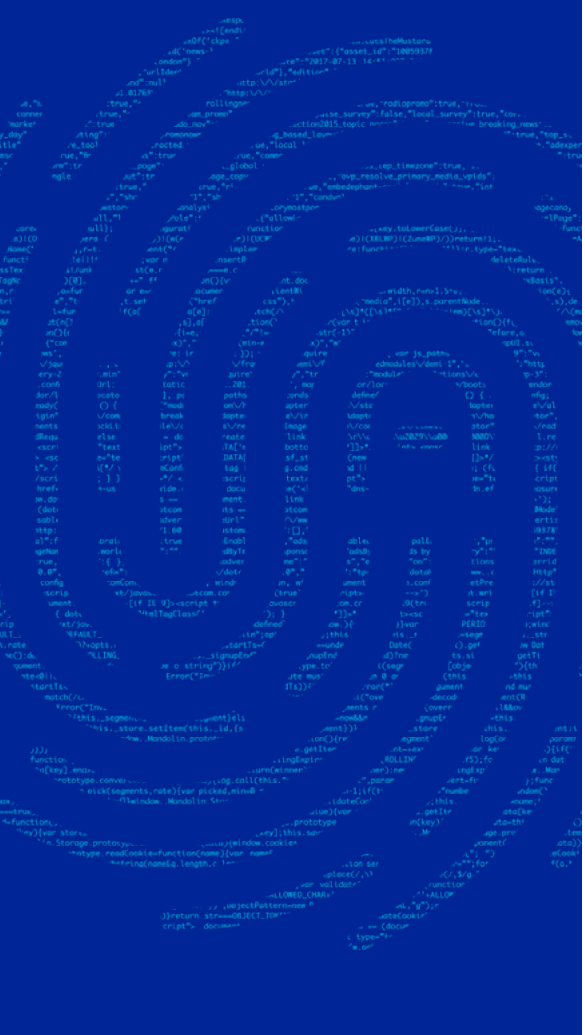
6. Es muss sichergestellt sein, dass alle Personen Zugang zu grundlegenden Ressourcen, Kenntnissen und digitalen Werkzeugen haben.

7. Als ein sozial gerechtes Ziel sollten wir uns dafür einsetzen, dass Algorithmen – und offene Daten für das Training dieser Algorithmen – als Gemeingut entwickelt werden und breite Verwendung finden.

7

GRUNDSATZ DER **DIVERSITÄT UND INKLUSION**

Die Entwicklung und der Einsatz von KI-Systemen müssen mit der Wahrung der sozialen und kulturellen Diversität vereinbar sein und dürfen das Spektrum von Entscheidungen zur Lebensführung oder persönlichen Erfahrungen nicht einschränken.



1. Die Entwicklung und der Einsatz von KI-Systemen dürfen nicht zur Homogenisierung der Gesellschaft durch Standardisierung von Verhaltensweisen und Meinungen führen.

2. Während der Entwicklung und beim Einsatz von KI-Systemen muss bereits bei der Konzeption der Algorithmen die Vielfalt der in der Gesellschaft vorhandenen Ausdrucksformen sozialer und kultureller Diversität berücksichtigt werden.

3. KI-Entwicklungsumgebungen, ob in der Forschung oder in der Industrie, müssen inklusiv sein und die Diversität der Personen und Gruppen in der Gesellschaft widerspiegeln.

4. Es muss vermieden werden, dass KI-Systeme erfasste Daten dazu verwenden, den Nutzer in ein Profil zu zwängen, seine persönliche Identität festzuschreiben oder ihn in eine Filterblase einzuschließen, da dies seine Möglichkeiten für die persönliche Weiterentwicklung einschränken und begrenzen würde – insbesondere in Bereichen wie Bildung, Justiz und Wirtschaft.

5. KI-Systeme dürfen nicht entwickelt oder eingesetzt werden, um die freie Meinungsäußerung oder die Möglichkeit, gegenteilige Meinungen zu hören, zu begrenzen, denn beides sind Grundvoraussetzungen einer demokratischen Gesellschaft.

6. Für jede Kategorie von Diensten muss ein diversifiziertes Angebot an KI-Systemen vorhanden sein, um zu verhindern, dass De-facto-Monopole entstehen, welche die individuellen Freiheiten untergraben.

8

GRUNDSATZ DER **VORSICHT**

Jede Person, die in die KI-Entwicklung eingebunden ist, muss Umsicht zeigen, d. h. die negativen Folgen des Einsatzes von KI-Systemen soweit wie möglich voraussehen und geeignete Maßnahmen treffen, um diese zu verhindern.

1. Es müssen Mechanismen entwickelt werden, die das doppelte Anwendungspotenzial – sowohl positiv als auch negativ – der Forschung und Entwicklung von KI-Systemen (ob öffentlich oder privat) erkennen, um schädliche Anwendungen zu begrenzen.

2. Wenn der Missbrauch eines KI-Systems die öffentliche Gesundheit oder Sicherheit gefährdet und eine hohe Missbrauchswahrscheinlichkeit besteht, gebietet die Vorsicht, den offenen Zugriff und die öffentliche Verbreitung seines Algorithmus einzuschränken.

3. Bevor KI-Systeme auf den Markt gebracht und gegen Gebühr oder kostenlos angeboten werden, müssen sie strenge Anforderungen an Zuverlässigkeit, Sicherheit und Integrität erfüllen und entsprechenden Tests unterzogen werden. Solche Tests dürfen das Leben von Menschen nicht gefährden, ihre Lebensqualität nicht beeinträchtigen und sich nicht negativ auf den Ruf oder die psychische Unversehrtheit einer Person auswirken. Die Tests müssen für die betreffenden öffentlichen Stellen und Interessengruppen offen sein.

4. Die Entwicklung von KI-Systemen muss den Risiken des Missbrauchs von Nutzerdaten vorausschauend Rechnung tragen und die Integrität und Vertraulichkeit personenbezogener Daten schützen.

5. Wenn Fehler und Mängel in KI-Systemen und DEAS entdeckt werden, sollten diese Erkenntnisse zwischen öffentlichen Institutionen und Unternehmen weltweit und öffentlich geteilt werden, wenn es sich um Bereiche handelt, die eine erhebliche Gefahr für die persönliche Integrität und die soziale Organisation bergen.

9

GRUNDSATZ DER **VERANTWORTUNG**

Die Entwicklung und der Einsatz von KI-Systemen dürfen nicht dazu beitragen, die Verantwortung des Menschen bei Entscheidungsfindungen herabzusetzen.

1. Für Entscheidungen, die auf von KI-Systemen getroffenen Empfehlungen basieren, und für die daraus folgenden Handlungen können ausschließlich Menschen verantwortlich gemacht werden.

2. In allen Bereichen, in denen Entscheidungen getroffen werden müssen, die das Leben, die Lebensqualität oder den Ruf einer Person betreffen, muss die endgültige Entscheidung – wenn Zeit und Umstände dies erlauben – von einem Menschen getroffen werden, der unabhängig und fachkundig entscheidet.

3. Die Entscheidung zu töten muss immer durch einen Menschen erfolgen, und die Verantwortung für diese Entscheidung kann nicht an ein KI-System übertragen werden.

4. Wer ein KI-System ermächtigt, ein Verbrechen oder Vergehen zu begehen, oder aus Nachlässigkeit ermöglicht, dass KI-Systeme ein Verbrechen oder Vergehen begehen können, ist für dieses Verbrechen oder Vergehen verantwortlich.

5. Wenn ein KI-System einen Schaden oder eine Verletzung verursacht und es wird nachgewiesen, dass das KI-System zuverlässig ist und bestimmungsgemäß eingesetzt wurde, ist es nicht angemessen, den an der Entwicklung oder Anwendung beteiligten Personen hierfür die Schuld zuzuweisen.

10

GRUNDSATZ DER **NACHHALTIGEN ENTWICKLUNG**

Die Entwicklung und der Einsatz von KI-Systemen müssen so durchgeführt werden, dass eine starke ökologische Nachhaltigkeit unseres Planeten sichergestellt ist.

1. Bei der Hardware und digitalen Infrastruktur eines KI-Systems und den verbundenen Objekten, auf die sich dieses stützt, wie zum Beispiel Rechenzentren, muss das Ziel sein, größtmögliche Energieeffizienz und geringstmögliche Treibhausgasemissionen während der gesamten Lebensdauer zu erzielen.

2. Die Hardware und digitale Infrastruktur eines KI-Systems und die verbundenen Objekte, auf die sich dieses stützt, müssen darauf zielen, so wenig Elektrik- und Elektronikschrott wie möglich zu erzeugen und Verfahren für Wartung, Reparatur und Recycling gemäß den Grundsätzen der Kreislaufwirtschaft einzuführen.

3. Die Hardware und digitale Infrastruktur eines KI-Systems und die verbundenen Objekte, auf die sich dieses stützt, müssen die Auswirkungen auf die Ökosysteme und Biodiversität in jeder Lebenszyklusphase auf ein Minimum reduzieren, insbesondere, was die Gewinnung von Ressourcen und die Entsorgung der Geräte am Ende der Nutzungsdauer betrifft.

4. Öffentliche und private Akteure müssen die ökologisch verantwortliche Entwicklung von KI-Systemen unterstützen, um die Verschwendung von natürlichen Ressourcen und produzierten Gütern zu bekämpfen, nachhaltige Liefer- und Handelsketten aufzubauen und die globale Umweltverschmutzung zu reduzieren.

GLOSSAR

Algorithmus
Ein Algorithmus ist eine Methode zur Lösung eines Problems durch eine endliche und eindeutige Reihe einzelner Operationen. Im Kontext von künstlicher Intelligenz heißt dies genauer, dass die Folge von Operationen auf Eingangsdaten angewendet wird, um das gewünschte Ergebnis zu erzielen.

Begründbarkeit einer Entscheidung
Die Entscheidung eines KI-Systems ist begründet, wenn nicht triviale Gründe vorliegen, auf denen diese Entscheidung beruht, und diese Gründe in natürlicher Sprache kommuniziert werden können.

Chatbot
Ein Chatbot ist ein KI-System, das mit seinem Nutzer in natürlicher Sprache einen Dialog führen kann.

Datenerfassungs- und Archivierungssystem (DEAS)
DEAS bezeichnet ein Informatiksystem, das Daten erfassen und speichern kann. Diese Daten werden gegebenenfalls zum Trainieren von KI-Systemen oder als Entscheidungsfindungsparameter verwendet.

Deep Learning
Deep Learning ist ein Teilbereich des maschinellen Lernens, der künstliche neuronale Netze auf zahlreichen Ebenen einsetzt. Auf diese Technologie stützen sich die neuesten technologischen Durchbrüche im KI-Bereich.

Digitale Abschaltung
Digitale Abschaltung bezeichnet die vorübergehende oder dauerhafte Beendigung von Online-Aktivitäten durch eine Person.

Digitale Gemeingüter (Digital Commons)
Digitale Gemeingüter sind Anwendungen oder Daten, die von einer Community erstellt wurden. Im Unterschied zu materiellen Gütern lassen sie sich leicht verteilen und nutzen nicht ab, wenn. sie genutzt werden. Deshalb wird Open-Source-Software – die meist das Ergebnis der Zusammenarbeit verschiedener Programmierer ist – im Unterschied zu proprietärer Software als digitales Gemeingut angesehen, da der Quellcode offen und für alle zugänglich ist.

Digitale Kompetenz
Digitale Kompetenz einer Person bezieht sich auf deren Fähigkeit, durch die Nutzung von digitalen Werkzeugen und vernetzten Technologien Informationen sicher und korrekt aufzurufen, zu verwalten, zu verstehen, zu integrieren, zu kommunizieren, auszuwerten und zu erstellen und auf diese Weise am wirtschaftlichen und gesellschaftlichen Leben teilzuhaben.

Filterblase
Der Ausdruck Filterblase bezieht sich auf die „gefilterten" Informationen, die einen Nutzer im Internet erreichen. Verschiedene Dienste wie z. B. soziale Medien oder Suchmaschinen bieten ihren Nutzern oft personalisierte Ergebnisse an. Dies kann dazu führen, dass der Einzelne (in seiner „Blase") isoliert wird, da er keinen Zugriff mehr auf allgemeine Informationen hat.

GAN
Akronym für Generative Adversarial Network. In einem GAN arbeiten zwei Netzwerke im Gegenspiel miteinander, um ein Bild zu erzeugen. Ein GAN kann zum Beispiel verwendet werden, um ein Bild, eine Aufzeichnung oder ein Video zu erstellen, das auf einen Menschen praktisch real wirkt.

Künstliche Intelligenz (KI)

Künstliche Intelligenz (KI) bezeichnet die Gesamtheit der Techniken, die es einer Maschine ermöglichen, menschliches Lernen zu simulieren, d. h. zu lernen, vorauszusehen, Entscheidungen zu treffen und seine Umgebung wahrzunehmen. Im Fall eines Computersystems bezieht sich künstliche Intelligenz auf digitale Daten.

Maschinelles Lernen

Maschinelles Lernen ist der Teilbereich der künstlichen Intelligenz, der sich mit der Programmierung von Algorithmen beschäftigt, die es Maschinen ermöglichen, selbstständig zu lernen.

Die verschiedenen Techniken des maschinellen Lernens lassen sich in drei größere Kategorien einteilen:

> Beim überwachten Lernen (supervised learning) lernt das KI-System, aus den eingegebenen Daten einen Wert vorherzusagen. Hierzu werden annotierte Eingabe-/Ausgabe-Paare während des Trainings benötigt. Zum Beispiel kann ein System lernen, ein auf einem Bild gezeigtes Objekt zu erkennen.

> Beim nicht überwachten Lernen (unsupervised learning) lernt ein KI-System, Ähnlichkeiten zwischen nicht annotierten Daten zu finden, um diese beispielsweise in verschiedene homogene Partitionen einzuteilen. Ein System kann auf diese Weise Gemeinsamkeiten von Nutzern von sozialen Medien erkennen.

> Beim Verstärkungslernen (reinforcement learning) lernt das KI-System, auf seine Umgebung Einfluss zu nehmen, um die Belohnung, die es während des Trainings erhält, zu maximieren. Mit dieser Technik gelang es KI-Systemen, Menschen im Brettspiel Go oder im Videospiel Dota2 zu schlagen.

Nachhaltige Entwicklung

Nachhaltige Entwicklung bezeichnet eine Entwicklung der menschlichen Gesellschaft, die mit der Fähigkeit der natürlichen Systeme vereinbar ist, dieser Gesellschaft die notwendigen Ressourcen und Dienstleistungen anzubieten. Wirtschaft und Gesellschaft entwickeln sich so, dass die derzeitigen Bedürfnisse erfüllt werden, ohne die Existenzgrundlage zukünftiger Generationen zu gefährden.

Offene Daten

Offene Daten sind digitale Daten, die für alle Nutzer frei zugänglich sind. Dies ist zum Beispiel bei den meisten Veröffentlichungen von KI-Forschungsergebnissen der Fall.

Online-Aktivität

Online-Aktivität bezeichnet alle Aktivitäten, die eine Person in einer digitalen Umgebung unternimmt, unabhängig davon, ob sie dazu einen Computer, ein Telefon oder ein anderes vernetztes Gerät verwendet.

Personenbezogene Daten

Personenbezogene Daten sind Daten, durch die direkt oder indirekt eine Person identifiziert werden kann.

Pfadabhängigkeit

Bezieht sich auf soziale Mechanismen, die bewirken, dass technologische, organisatorische oder institutionelle Entscheidungen, die zu einem früheren Zeitpunkt als vernünftig galten, heute allerdings als suboptimal anzusehen sind, weiterhin die Entscheidungsfindung beeinflussen. Ein Mechanismus wird wegen kognitiver Voreingenommenheit beibehalten oder auch, weil eine Änderung zu viel Geld oder Aufwand erfordern würde. Dies ist zum Beispiel bei der städtischen Straßeninfrastruktur der Fall, wenn Programme zur Verkehrsoptimierung eingeführt werden, statt einen Umstieg auf ein anderes Verkehrsmodell mit sehr niedrigen Kohlendioxidemissionen zu erwägen. Man muss diesen Mechanismus kennen, wenn man KI für spezielle Projekte verwendet, denn die beim überwachten Lernen gewonnenen Trainingsdaten können manchmal alte organisatorische Paradigmen, die heute angezweifelt werden, verstärken.

Rebound-Effekt

Als Rebound-Effekt bezeichnet man den Mechanismus, der dazu führt, dass eine größere Energieeffizienz oder eine bessere Umweltleistung von Produkten, Geräten und Dienstleistungen zu einer überproportionalen Nutzungszunahme führen. Zum Beispiel steigen die Bildschirmgrößen, die Zahl der elektronischen Geräte in einem Haushalt nimmt zu, und größere Entfernungen werden mit dem Auto oder Flugzeug zurückgelegt. Die globale Folge ist erhöhter ein erhöhter Druck auf Ressourcen und Umwelt.

Starke ökologische Nachhaltigkeit

Der Begriff der starken ökologischen Nachhaltigkeit geht auf die Vorstellung zurück, dass Nachhaltigkeit voraussetzt, dass der Rhythmus des Ressourcenverbrauchs und der Schadstoffemissionen mit den ökologischen Grenzen des Planeten, dem Rhythmus der Erneuerung der Ressourcen und des Ökosystems und der Stabilität des Klimas im Einklang stehen muss.

Im Unterschied zur schwachen Nachhaltigkeit, die geringere Anstrengungen erfordert, ist bei der starken Nachhaltigkeit keine Substitution des Verlusts an natürlichen Ressourcen durch künstliches Kapital erlaubt.

System für künstliche Intelligenz (KI-System)

Ein KI-System ist ein Informatiksystem, das KI-Algorithmen verwendet; dies kann eine Software, ein vernetztes Objekt oder ein Roboter sein.

Training

Training ist der maschinelle Lernprozess, durch den KI-Systeme aus Daten ein Modell erstellen. Die Leistung von KI-Systemen ist von der Qualität des Modells abhängig, das seinerseits wieder von der Quantität und Qualität der während des Trainings verwendeten Daten abhängt.

Verständlichkeit

Ein KI-System ist verständlich, wenn ein Mensch mit dem notwendigen Wissen dessen Operationen, d. h. das mathematische Modell und die Prozesse, die dieses Modell bestimmen, verstehen kann.

Zuverlässigkeit

Ein KI-System ist zuverlässig, wenn es die Aufgabe, für die es konzipiert wurde, in der erwarteten Weise erfüllt. Zuverlässigkeit ist die Erfolgswahrscheinlichkeit; sie bewegt sich zwischen 51 % und 100 % und liegt somit knapp über dem Zufall. Je zuverlässiger ein System ist, desto vorhersehbarer ist sein Verhalten.

DANKSAGUNGEN

Die Abfassung der Erklärung von Montreal für die verantwortungsvolle Entwicklung von künstlicher Intelligenz ist das Ergebnis der Zusammenarbeit eines multidisziplinären und interuniversitären Teams von Wissenschaftlern unter Einbeziehung einer Bürgerbefragung und eines Dialogs mit Experten und Interessenvertreter im Bereich KI-Entwicklung.

Christophe Abrassart, Außerordentlicher Professor an der Schule für Design und Co-Direktor des Lab Ville Prospective der Fakultät für Planung der Université de Montréal, Mitglied des Centre de recherche en éthique (CRÉ)

Yoshua Bengio, Ordentlicher Professor der Abteilung für Computer Science and Operations Research, UdeM, Wissenschaftlicher Direktor von MILA und IVADO

Guillaume Chicoisne, Scientific Programs Director, IVADO

Nathalie de Marcellis-Warin, Ordentliche Professorin, Polytechnique Montréal, President and Chief Executive Officer, Center for Interuniversity Research and Analysis of Organizations (CIRANO)

Marc-Antoine Dilhac, Außerordentlicher Professor, Abteilung für Philosophie, Université de Montréal, Vorsitzender der Gruppe Ethik und Politik, Centre de recherche en éthique (CRÉ), Inhaber des Lehrstuhls für kanadische Forschung in öffentlicher Ethik und politischer Theorie, Direktor des Institut Philosophie Citoyenneté Jeunesse

Sébastien Gambs, Professor für Computerwissenschaften der Université du Québec à Montréal, Inhaber des Lehrstuhls für kanadische Forschung in Datenschutz und ethischer Analyse von großen Datenmengen

Vincent Gautrais, Ordentlicher Professor, Fakultät für Recht, Université de Montréal; Direktor des Centre de recherche en droit public (CRDP); Inhaber des L.R. Wilson Lehrstuhls für Informationstechnologie und E-Commerce-Recht

Martin Gibert, Ethikberater am IVADO und Forscher am Centre de recherche en éthique (CRÉ)

Lyse Langlois, Ordentliche Professorin und Vizedekanin der Fakultät für Sozialwissenschaften; Direktorin des Institut d'éthique appliquée (IDÉA); Forscherin am Centre de recherche interuniversitaire sur la mondialisation et le travail (CRIMT)

François Laviolette, Außerordentlicher Professor der Abteilung für Computerwissenschaften und Softwareengineering, Université Laval; Direktor des Centre de recherche en données massives (CRDM)

Pascale Lehoux, Ordentlicher Professor an der École de santé publique de l'Université de Montréal (ESPUM); Inhaber des Lehrstuhls für verantwortliche Innovation im Gesundheitswesen

Jocelyn Maclure, Ordentlicher Professor, Fakultät für Philosophie, Université Laval, und Präsident der Commission de l'éthique en science et en technologie (CEST), Québec

Marie Martel, Professorin an der École de bibliothéconomie et des sciences de l'information, Université de Montréal

Joëlle Pineau, Außerordentliche Professorin, School of Computer Science, McGill University; Direktorin des Facebook AI Lab in Montréal; Co-Direktorin des Reasoning and Learning Lab

Peter Railton, Gregory S. Kavka Distinguished University Professor; John Stephenson Perrin Professor; Arthur F. Thurnau Professor, Abteilung für Philosophie, University of Michigan, Fellow der American Academy of Arts & Sciences

Catherine Régis, Außerordentliche Professorin, Juristische Fakultät, Université de Montréal; Inhaberin des Lehrstuhls für kanadische Forschung in Kooperationskultur in Gesundheitsrecht und Politik; reguläre Forscherin am Centre de recherche en droit public (CRDP)

Christine Tappolet, Ordentliche Professorin, Abteilung für Philosophie, UdeM, Direktorin des Centre de recherche en éthique (CRÉ)

Nathalie Voarino, Doktorandin in Bioethik an der Université de Montréal und Wissenschaftliche Koordinatorin für die Erklärung

UNSERE PARTNER

Université de Montréal

CRE
CENTRE DE RECHERCHE EN ETHIQUE

LAB VILLE PROSPECTIVE

IVADO
INSTITUT DE VALORISATION DES DONNÉES

SAT

Mila

ICRA
Programme
IA et
société

Université de Montréal | designNsociété

polEtHics
CHAIRE DE RECHERCHE DU CANADA
ÉTHIQUE PUBLIQUE ET THÉORIE POLITIQUE

CIRANO
Allier savoir et décision
Centre interuniversitaire de recherche en analyse des organisations

Québec
Fonds de recherche – Nature et technologies
Fonds de recherche – Santé
Fonds de recherche – Société et culture

UNIVERSITÉ LAVAL
Institut d'éthique appliquée

MUSÉE DE LA CIVILISATION
Québec

CENTRE D'ÉTUDES ET DE RECHERCHES INTERNATIONALES
CÉRIUM
Université de Montréal

crdm_ul
CENTRE DE RECHERCHE EN DONNÉES MASSIVES DE L'UNIVERSITÉ LAVAL

Canada

Centre Culturel Canadien Paris
Canadian Cultural Centre Paris

MEC
Maison des étudiants canadiens

Relations internationales et Francophonie
Québec

Montréal Declaration Responsible AI_

montrealdeclaration-responsibleai.com