



## Schlusserklärung



### Schlusserklärung der Konferenz über

#### Robotik, KI und Humanität, Wissenschaft, Ethik und Politik

veranstaltet von der Päpstlichen Akademie der Wissenschaften (PAS) und der Päpstlichen Akademie der Sozialwissenschaften (PASS), 16. - 17. Mai 2019 in der Casina Pio IV, Vatikan-Stadt.

#### THEMEN UND AGENDA

1. Die Fortschritte im Bereich des Maschinellen Lernens (oft auch als Künstliche Intelligenz (KI) bezeichnet) und der Robotik beschleunigen sich. Sie haben erhebliche Auswirkungen auf das Funktionieren von Gesellschaften und Volkswirtschaften und haben zu einer breiten Debatte über ihre Vor- und Nachteile für die Menschheit geführt. **Dieses sich schnell entwickelnde Gebiet der Wissenschaft und Technologie erfordert unsere sorgfältige Aufmerksamkeit.** Die Päpstlichen Akademien haben daher diese Konferenz organisiert und Kollegen aus verschiedenen Disziplinen zusammengebracht [1].
2. Künstliche Intelligenz und Anwendungen in Robotertechnologien werden **weitreichende Auswirkungen auf Wirtschaft, Gesellschaft und Kultur** haben. Zusätzlich zur Untersuchung der derzeitigen Forschungshorizonte in der KI/Robotik haben wir die vermutlichen Auswirkungen auf das gesellschaftliche Wohlergehen sowie die ethischen und politischen Implikationen überprüft und diskutiert.
3. KI und Robotik sind vielversprechend, um einige unserer schwierigsten sozialen, wirtschaftlichen und ökologischen Probleme wie den **Klimawandel und die Erreichung der UN-Ziele** für nachhaltige Entwicklung (SDGs) für 2030 anzugehen.
4. Es gibt auch langfristige Trends in der KI und Robotik, mit Folgen, die letztlich den **Platz des Menschen in der Gesellschaft** herausfordern können.
5. Wachsende Sorge bereiten die Risiken für den Frieden durch **neue Formen der Kriegsführung** (Cyberangriffe, autonome Waffen), die neue internationale Sicherheitsbestimmungen erfordern.
6. **Ethische und religiöse Aspekte** von KI und Robotik müssen geklärt werden, um den potenziellen Bedarf an regulatorischen Maßnahmen für Anwendungen und die zukünftige Entwicklung von KI/Robotik zu steuern.

#### GRUNDLEGENDE FRAGEN IN DER KI UND ROBOTIK

7. Im Bereich der KI hat sich eine Vielzahl von theoretischen Ansätzen und Frameworks auf der einen Seite und immer weitreichendere praktische Anwendungen auf der anderen Seite entwickelt. **KI hat das Potenzial, Fortschritte in allen Bereichen von Wissenschaft und Gesellschaft zu erzielen.** Sie kann uns helfen, einige unserer kognitiven Einschränkungen zu überwinden und komplexe Probleme zu lösen. Während riesige Datenmengen eine Herausforderung für die menschlichen kognitiven Fähigkeiten darstellen, bieten Big Data beispiellose Möglichkeiten für die Wissenschaft. Das Translationspotenzial von Big Data ist beträchtlich, zum

Beispiel in den Bereichen **Medizin, Gesundheitswesen, Bildung sowie Management komplexer Systeme im Allgemeinen (Biosphäre, Geosphäre, Wirtschaft)**. Die auf großen Daten basierende Wissenschaft als solche bleibt jedoch empirisch und fordert uns heraus, die zugrundeliegenden kausalen Mechanismen zu entdecken, die Muster erzeugen.

8. In Kombination mit Robotik und Gehirn-Computer-Schnittstellen bietet die KI bereits heute eine einzigartige **Unterstützung für Patienten mit sensorischen oder motorischen Defiziten und erleichtert die Betreuung behinderter Patienten**. Durch die Bereitstellung neuer Instrumente für den Wissenserwerb kann die KI dramatische Veränderungen in der Bildung bewirken und den Zugang zu Wissen erleichtern. Synergien sind auch möglich, die sich aus Roboter-zu-Roboter-Interaktionen ergeben, sowie potenzielle positive Synergien zwischen Mensch und Roboter, die aus der Zusammenarbeit an Aufgaben entstehen.

9. Zu den grundlegenden Belangen der KI und Robotik gehört **die Frage, ob Maschinen hypothetisch Fähigkeiten wie Bewusstsein erlangen können**. Dieser Kernpunkt wird derzeit aus gegensätzlichen Perspektiven von Naturwissenschaft, Gesellschaftstheorie und Philosophie diskutiert; dies bleibt ein offenes Thema, zumal es eine Vielfalt von Definitionen von "Bewusstsein" gibt. Es stellt sich die Frage, ob die Betonung der übermenschlichen Kapazitäten der KI für Berechnung und Kompilierung die vielen Grenzen künstlicher Systeme verdeckt.

10. Die meisten behaupten, dass **Roboter nicht als Personen** betrachtet werden können, also werden und sollten Roboter keine freien Agenten sein oder Rechte besitzen. Einige argumentieren jedoch, dass "Command and Control"-Konzepte für Mensch-Roboter Beziehungen möglicherweise nicht angemessen sind und andere erwägen sogar, ob etwas wie "elektronische Staatsbürgerschaft" in Betracht gezogen werden kann. Die christliche Philosophie und Theologie besagt, dass die menschliche Seele von Natur aus unzerstörbar ist. Dies ist die metaphysische Grundlage, nach der der Mensch in sich selbst frei und zur ethischen Ordnung fähig ist und aus den Kräften der Natur hervorgeht. Als spirituelles Subjekt ist der Mensch Imago Dei. In diesem Sinne der christlichen Philosophie können KI/Roboter nicht als Personen betrachtet werden, so dass **Roboter keine menschliche Freiheit und keine geistliche Seele besitzen**. Sie sollten nicht als "Bilder Gottes", sondern vielleicht **als "Bilder des Menschen" zu betrachten** sein, die vom Menschen als ihre Werkzeuge zum Wohle der menschlichen Gesellschaft geschaffen werden.

11. Innerhalb der maschinellen Lernforschung gibt es eine Entwicklungslinie, die darauf abzielt, theoretische Grundlagen für das **Design kognitiver Agenten** zu identifizieren; solche Grundlagen würden die Ableitung von Theoremen ermöglichen, die die Möglichkeiten und Grenzen intelligenter Agenten charakterisieren. Kognitive Agenten agieren in einem offenen, teilweise oder völlig unbekannten Umfeld, um Ziele zu erreichen. **Schlüsselkonzepte für ein grundlegendes Framework für KI und Robotik** umfassen demnach: Agenten, Umgebungen, Belohnungen, lokale Scores, globale Scores, das genaue Modell der Interaktion zwischen Agenten und Umgebungen und eine Spezifikation der verfügbaren Rechenressourcen von Agenten und Umgebungen. Ein "intelligenter Agent" kann definiert werden als ein Agent, der in einer Vielzahl von Umgebungen Ziele erreichen kann. Das Thema Kognition mit begrenzten Ressourcen verdient weitere Untersuchungen, um einen sinnvollen und breit integrierten fundamentalen Rahmen für künstliche Intelligenz zu erreichen.

12. Ein zentraler Aspekt des Lernens aus der Erfahrung ist die **Darstellung und Verarbeitung von unsicherem Wissen**. In Ermangelung deterministischer Annahmen über die Welt gibt es keine nichttriviale logische Schlussfolgerung, die aus der Vergangenheit für jedes zukünftige Ereignis gezogen werden kann. Dementsprechend ist es von grundlegendem Interesse, die Struktur der Unsicherheit als eigenständige Frage zu analysieren. Es bleibt noch viel zu forschen, um vorläufige Vorschläge und formale Methoden aus dem theoretischen Bereich in das Engineering effizienter algorithmischer Lösungen umzusetzen.

13. Einige neuere Ergebnisse stellen einen engen Zusammenhang zwischen Lernfähigkeit und Beweisbarkeit her und reduzieren damit die Frage, was effektiv gelernt werden kann, auf die grundlegenden Fragen der Mathematik im Hinblick auf Existenzaxiome der Mengenlehre. Ergebnisse der inversen Mathematik, ein Zweig der mathematischen Logik, der Theoreme mit Bezug auf die zu ihrem Nachweis notwendigen Mengen-Existenz-Axiome analysiert, können verwendet werden, um die Auswirkungen formaler Frameworks für maschinelles Lernen zu veranschaulichen. Im Allgemeinen werden **Modellprüfungs- und Proofchecking-Techniken** immer wichtiger, je mehr die Kritikalität der Aufgaben, die intelligenten Agenten übertragen werden, zunimmt.

14. Bis vor kurzem hatte die mathematische Grundlagenforschung nur wenige (wenn überhaupt) ethische Fragen auf der Tagesordnung. Da Mathematiker und Software-Designer jedoch für die Entwicklung von KI von zentraler Bedeutung sind, ist es unerlässlich, dass sie die ethischen Auswirkungen ihrer Arbeit berücksichtigen. Angesichts der immer häufiger gestellten Fragen zur Vertrauenswürdigkeit autonomer Systeme haben **KI-**

**Entwickler die Verantwortung - die idealerweise zu einer gesetzlichen Verpflichtung werden sollte -, vertrauenswürdige und steuerbare Robotersysteme zu schaffen.**

## **WISSENSCHAFT UND ENGINEERING VON MENSCH-KI/ROBOTER INTERAKTIONEN**

15. Zu den Domänen, die uns als Menschen definieren, und in der KI intensiv beforscht werden, gehören **Sprache, Symbolverarbeitung, One-Shot-Lernen, Selbsteinschätzung, Selbstvertrauen**, Programminduktion, Zielkonzeption und die Zusammenführung bestehender Module in eine übergreifende, multifunktionale intelligente Architektur. Computational Agents, die durch Verstärkungslernen und **Deep Learning Frameworks** geschult wurden, zeigen hervorragende Leistungen bei Aufgaben, die bisher als unlösbar galten. Während eine vollständige Fundierung für eine allgemeine Theorie der computergestützten kognitiven Agenten noch fehlt, hat der konzeptionelle und praktische Fortschritt der künstlichen Intelligenz einen Zustand erreicht, in dem Ethik- und Sicherheitsfragen und die Auswirkungen auf die Gesellschaft insgesamt zu drängenden Themen werden. Ein solches Problem stellt sich beispielsweise, wenn KI-basierte Rückschlüsse auf die Gefühle von Personen aus deren Gesichtserkennungsdaten abgeleitet werden.

16. Die Verbreitung der Robotik verändert die menschlichen und sozialen Beziehungen in allen Bereichen der Gesellschaft, in der Familie sowie am Arbeitsplatz und in der Öffentlichkeit grundlegend. Diese Modifikationen nehmen den Charakter von **Hybridisierungsprozessen an zwischen den eigentlichen menschlichen Merkmalen von Beziehungen und den künstlichen, also zwischen analoger und virtueller Realität**. Daher ist es notwendig, die wissenschaftliche Forschung zu Fragen der sozialen Auswirkungen zu verstärken, die sich aus der Übertragung relevanter Aspekte sozialer Organisation an KI und Roboter ergeben. Ein Ziel dieser Forschung sollte es sein zu verstehen, wie es möglich ist, die relevanten Veränderungsprozesse zu steuern und jene Beziehungsgüter zu produzieren, die eine vorteilhafte individuelle Erfüllung innerhalb einer nachhaltigen und fairen gesellschaftlichen Entwicklung realisieren.

17. Wir stellen fest, dass die schnellen Fortschritte in der Robotik ganze Branchen verändern (Industrie 4.0). Die Entwicklung des Internets der Dinge (IoT) mit der Kommunikation zwischen Maschinen und dem vernetzten maschinellen Lernen bietet große Fortschritte für Dienstleistungen wie Bank- und Finanzdienstleistungen. Roboter-Roboter- und Mensch-Roboter-Interaktionen werden immer intensiver. Dennoch sind **KI-Systeme schwer zu testen und zu validieren**. Das macht das Vertrauen in KI und Roboter herausfordernd. Fragen der Regulierung und des Eigentums an Daten, der Zuweisung von Verantwortlichkeiten und der Transparenz von Algorithmen stellen sich und erfordern legitime institutionelle Regelungen.

18. Wir können zwischen **mechanischen Robotern**, die für Routineaufgaben in der Produktion entwickelt wurden, und den **Fähigkeiten der KI zur Unterstützung** der sozialen Versorgung, medizinischer Verfahren, sicherer und energieeffizienter Mobilitätssysteme, Bildungsaufgaben und wissenschaftlicher Forschung unterscheiden. Während intelligente Assistenten sowohl Erwachsenen als auch Kindern zu Gute kommen können, tragen sie auch Risiken, weil ihre Auswirkungen auf das sich entwickelnde Gehirn unbekannt sind und weil Menschen die Motivation in Bereichen verlieren können, in denen die KI überlegen erscheint.

19. In den letzten Jahrzehnten hat der Bereich der Robotik eine **Vielzahl von Anwendungen für neuartige Dienstleistungen und Unterstützungen** hervorgebracht. Paradigmatisch für viele Anwendungsszenarien sind Roboter-Handarmsysteme, bei denen die Herausforderungen an Präzision, Sensibilität und Robustheit mit sicheren Greifanforderungen einhergehen. Obwohl Roboterhände noch weit von ihren menschlichen Vorbildern entfernt sind, wurde ihre Leistung durch neue Steuerungsmethoden erheblich gesteigert. Vielversprechende Anwendungen entwickeln sich für Telemanipulationssysteme in einer Vielzahl von Bereichen wie Gesundheitswesen, Fabrikproduktion und Mobilität.

20. Die **künstliche Intelligenz kann einer verantwortungsvollen sozialen Steuerung und Koordination dienen**, einschließlich der Identifizierung und Prävention illegaler Transaktionen, z.B. von Geld, das aus kriminellen Aktivitäten wie Drogenhandel, Menschenhandel oder illegalen Transplantationen stammt. Wenn sich die KI jedoch allein in den Händen von Unternehmen befindet, werden die aus der KI geschöpften Gewinne kaum gerecht verteilt werden. **Diese neuen Technologien dürfen nicht zu Instrumenten werden, um Menschen zu versklaven oder die Armen weiter zu marginalisieren.**

## **ROBOTIK, DIE DIE ZUKUNFT VON ARBEIT, LANDWIRTSCHAFT, ARMUT, UNGLEICHHEIT UND ÖKOLOGIE VERÄNDERT**

21. Wir haben KI-Anwendungen (und damit zusammenhängende neue Technologien) in den Bereichen **Medizin und Gesundheitswesen, Mobilität und Verkehr, Fertigung und Landwirtschaft** untersucht. Es wurden große Chancen festgestellt und den Robotik / KI-Anwendungen in jedem dieser Bereiche wurde entsprechend große Aufmerksamkeit geschenkt. Eine sektorale Perspektive auf KI und Robotik hat jedoch Grenzen. Es ist **notwendig, sich ein umfassenderes Bild von den Zusammenhängen zwischen den**

**Anwendungen zu machen** und sich auf Politiken zu konzentrieren, die eine allgemeine Fairness und Gerechtigkeit ermöglichen, die alle Aspekte von KI und Robotik abdeckt.

22. Wenn die KI nicht für den öffentlichen Nutzen kanalisiert wird, kann sie große Probleme für die Wirtschaft und die Stabilität der Gesellschaft aufwerfen. **Arbeitsplätze können durch computergestützte Geräte verloren gehen, was zu einer Zunahme der Einkommensungleichheit und zu einer Wissenserosion führen kann.** Fortschritte in der Automatisierung und ein erhöhtes Angebot an künstlicher Arbeitskraft, insbesondere in der Landwirtschaft und in der Industrie, können die **Beschäftigung in den Entwicklungs- und Schwellenländern** deutlich reduzieren. Durch Verknüpfungen innerhalb globaler Wertschöpfungsketten können Arbeitnehmer in Ländern mit niedrigem Einkommen auch von Robotern in Ländern mit höherem Einkommen betroffen sein, was den Bedarf an der Auslagerung von Routinearbeiten in die ehemaligen Niedriglohnregionen reduzieren könnte. Der Einsatz von Robotern könnte jedoch auch den Bedarf an Arbeitskräften erhöhen, indem er die Produktionskosten senkt, was zu einer industriellen Expansion führt. Zuverlässige Schätzungen über neue Arbeitsplätze in den Branchen Design und Herstellung von Robotern fehlen, sind aber notwendig.

23. Die Auswirkungen der Robotik auf Beschäftigung und Arbeit sind ein wichtiges politisches Thema. Politik sollte darauf abzielen, die notwendigen Maßnahmen der sozialen Sicherheit für die betroffenen Arbeitnehmer bereitzustellen und gleichzeitig in die Entwicklung der notwendigen Fähigkeiten zu investieren, um die neu geschaffenen Arbeitsplätze zu nutzen. Der Staat sollte in der Lage sein, **Gewinne, die durch die Arbeit der Robotik erzielt werden, umzuverteilen.** Eine solche Umverteilung könnte beispielsweise die Umschulung der Betroffenen finanzieren, damit sie in Beschäftigung bleiben können. In diesem Zusammenhang ist es wichtig, sich daran zu erinnern, dass viele dieser neuen technologischen Innovationen mit Unterstützung durch öffentliche Mittel realisiert wurden.

24. Roboter, KI und digitales Kapital im Allgemeinen können als Steuerbasis betrachtet werden. Derzeit ist dies nicht der Fall, sondern **menschliche Arbeit wird besteuert, aber Roboterarbeit nicht.** Auf diese Weise werden Robotersysteme indirekt subventioniert, da Unternehmen sie in ihren Buchhaltungssystemen abschreiben können, was die Unternehmensbesteuerung reduziert. Diese **Verzerrungen, die menschliche Arbeitskräfte benachteiligen und gleichzeitig Investitionen in Roboter begünstigen, sollten korrigiert werden.**

25. Wir stellen fest, dass **die Auswirkungen, Chancen und Risiken von KI und Robotik für eine nachhaltige Entwicklung und für die Armen mehr Aufmerksamkeit erfordern.** Insbesondere die Auswirkungen auf Länder mit niedrigem Einkommen, auf Marginalisierte und auf Frauen müssen untersucht und in Programmen und Richtlinien berücksichtigt werden. KI-Lehrmittel in vielen Regionen mit niedrigem Einkommen sind eine Chance. Da ein großer Teil der Armen auf kleinen Farmen lebt, insbesondere in Afrika sowie in Süd- und Ostasien, ist es wichtig, dass sie Zugang zu sinnvollen digitalen Technologien und KI erhalten. Beispiele sind die Zertifizierung von Grundeigentum durch Blockchain-Technologie, Präzisionstechnologien im Land- und Pflanzenmanagement und vieles mehr.

26. **Direkte und indirekte Umweltauswirkungen von KI / Robotik sollten stärker berücksichtigt werden.** Die Überwachung durch intelligente Fernerkundung in terrestrischen und aquatischen Systemen kann erheblich verbessert werden, um Veränderungen der biologischen Vielfalt und die Auswirkungen von Maßnahmen zu bewerten. Es gibt aber auch das Problem der Verschmutzung durch Elektroschrott, der von Industrieländern in Ländern mit niedrigem Einkommen entsorgt wird. Dies muss beendet werden, und der CO<sub>2</sub>-Fußabdruck von KI und Robotik erfordert auch mehr Aufmerksamkeit.

## **ROBOTIK, KI UND MILITÄRISCHE KONFLIKTE**

27. Innerhalb **militärischer Konflikte können KI-basierte Systeme (einschließlich Roboter) einer Vielzahl von Zwecken dienen,** unter anderem der Extraktion von verwundeten Personen, der Überwachung der Einhaltung von Kriegsrecht und Einsatzregeln, der Verbesserung des situativen Bewusstseins, der Gefechtsfeldplanung und des Treffens von strategischen und taktischen Entscheidungen. Dies wirft anspruchsvollste moralische Fragen auf. In jedem Fall müssen die Auswirkungen **potenziell gesenkter Kriegsbarrieren und Systemrisiken** sorgfältig geprüft werden, bevor eine Umsetzung auf dem Gefechtsfeld erwogen werden kann.

28. Die Sorge, im **Wettlauf um die Entwicklung neuer militärischer Anwendungen** ins Hintertreffen zu geraten, darf nicht zu einer Entschuldigung für Fehlschlüsse in der Sicherheitsforschung, bei Tests und einer angemessenen Ausbildung werden. Da der Trend im Waffendesign eher weg von einer groß angelegten Infrastruktur zu autonomen, dezentralen und miniaturisierten Systemen weist, werden die Risiken zerstörerischer Effekte aufgrund von Systemdesigns und potenziellem Versagen in Bezug auf die meisten der heutigen operativen Systeme stark zunehmen. Dieses erhöhte Potenzial für negative externe Effekte muss

durch entsprechende Investitionen in Sicherheit und Ausbildung ausgeglichen werden. KI-Tools sollten die Ausübung eines soliden und moralischen Urteils durch Militärpersonal verbessern und nicht schmälern.

29. Konzertierte internationale Bemühungen müssen darauf ausgerichtet sein, die spezifischen Anwendungen der KI zu identifizieren, die ein Eskalationsrisiko darstellen. **Die Staaten sollten sich auf konkrete Schritte einigen, um das Risiko von KI-gestützten und möglicherweise eskalierten Kriegen zu verringern.** Dabei sollte nicht nur auf die Gefahren der Technologie als Ersatz für Menschen in militärischen Bereichen geachtet werden, sondern auch auf die Gefahr, dass die KI für die Ausübung der "strategischen Reflexion" in Konfliktsituationen eine Rolle spielt.

30. In Bezug auf tödliche **autonome Waffensysteme** sollten nach dem derzeitigen Stand der technischen Kompetenz (und auf absehbare Zeit) keine Systeme eingesetzt werden, die im unbeaufsichtigten Modus funktionieren. Die **menschliche Rechenschaftspflicht** muss aufrechterhalten werden, damit die Einhaltung international anerkannten Kriegsrechts gewährleistet ist und Verstöße geahndet werden können.

## **GESELLSCHAFT, ETHISCHE, RELIGIÖSE UND REGULATORISCHE ASPEKTE DER ROBOTIK/AI**

Zusätzlich zu den bereits angesprochenen Punkten sollten die folgenden Belange hervorgehoben werden:

31. Die Anstrengungen einer öffentlich geförderten Entwicklung intelligenter Maschinen sollten auf das Gemeinwohl ausgerichtet sein. Die **Verbesserungen öffentlicher Güter und Dienstleistungen sowie Gesundheit, Bildung, Glück und Nachhaltigkeit müssen von vorrangig sein.** KI kann unerwartete Verzerrungen oder unmenschliche Folgen haben, einschließlich der gesellschaftlichen Segmentierung und der Diskriminierung nach Rasse und Geschlecht, und diese müssen angegangen werden, bevor sie auftreten können. Dies sind nationale und globale Fragen, und letztere bedürfen einer weiteren **Aufmerksamkeit der Vereinten Nationen.**

32. Was die Privatsphäre, den Zugang zu neuem Wissen und die Informationsrechte betrifft, so sind **die Armen aufgrund ihres Mangels an Macht und Stimme besonders gefährdet.** KI und Robotik müssen von einer stärkeren Befähigung der Armen durch Information und Bildung begleitet werden und die berufliche Qualifikation müssen verbessert werden.

33. **Das Thema Arbeit** steht im Mittelpunkt der Soziallehre der Kirche. Wie in *Laborem exercens* festgestellt, ist der Zugang zu sinnvoller Arbeit eine grundlegende Voraussetzung für die Menschenwürde. Die Beseitigung dieses Zugangs durch den Einsatz von Maschinen ist nicht akzeptabel, während die Verringerung der Arbeitsbelastung und der gesundheitlichen Risiken der Arbeit eine Entwicklung darstellt, die zu begrüßen ist.

34. Die Politik sollte darauf abzielen, die Vorteile der Produktivitätszunahme zu teilen und zwar durch eine Kombination aus Gewinnbeteiligung, indem **statt einer Subventionierung von Robotern eine (digitale) Kapitalbesteuerung** in Betracht gezogen wird, sowie durch eine Verringerung der Arbeitszeit für Routineaufgaben.

35. Die **Risiken manipulativer Anwendungen von KI** zur Meinungsbildung und Wahlintervention erfordern Aufmerksamkeit, und hier sind nationale und internationale Kontrollen erforderlich.

36. KI und Robotik bieten große Chancen und bergen Risiken; daher sollten **Regulierungen durch legitime öffentliche Einrichtungen angemessen gestaltet werden**, ohne die Chancen zu beeinträchtigen und ohne übermäßige Risikobereitschaft und Verzerrung zu fördern. Dazu bedarf es eines Rahmens, in dem der integrative öffentliche gesellschaftliche Diskurs von Wissenschaftlern verschiedener Disziplinen geprägt wird und alle Teile der Gesellschaft teilnehmen können.

37. Es werden **neue Formen der Regulierung der digitalen Wirtschaft erforderlich**, die einen angemessenen Datenschutz und die Privatsphäre gewährleisten. Darüber hinaus müssen deontische Werte wie "erlaubt", "obligatorisch", "verboten" prägend gefestigt werden, um im Web zu navigieren und mit Robotern zu interagieren.

38. **Unternehmen sollten Ethik- und Sicherheitsgremien einrichten** und sich mit gemeinnützigen Organisationen zusammentun, die best practices und Standards für den sinnvollen Einsatz von KI und Robotik etablieren wollen. Entsprechende Protokolle für die Sicherheit der KI-Roboter müssen entwickelt werden, wie z.B. die doppelte Überprüfung eines KI-Robotersystems durch unabhängige Konstruktionsteams. Das Bestehen von Ethik- und Sicherheitstests, die beispielsweise die sozialen Auswirkungen bewerten oder verdeckte rassistische Vorurteile identifizieren, sollte zu einer Voraussetzung für die Veröffentlichung neuer KI-Software werden. Externe zivile Gremien, die eine regelmäßige und transparente Bewertung aller Technologien, einschließlich des Militärs, durchführen, sollten in Betracht gezogen werden.

39. **Wissenschaftler und Ingenieure tragen als Entwickler von KI- und Robotersystemen Verantwortung** dafür, dass ihre Erfindungen und Innovationen sicher sind und für gute Zwecke genutzt werden können.

---

[1] Eine frühere Konferenz (30. November-1. Dezember 2016) wurde von den beiden Akademien über die Macht und die Grenzen der Künstlichen Intelligenz organisiert. Die vorliegende Erklärung baut auf der Arbeit dieser früheren Konferenz auf (einschließlich der Erklärung, die zu ihrem Abschluss abgegeben wurde).

Statement verfasst von Joachim von Braun (PAS-Präsident) mit signifikanten Beiträgen der Konferenzteilnehmer

### **Konferenzveranstalter**

von Braun, Joachim | PAS Präsident, Universität Bonn, Zentrum für Entwicklungsforschung, (ZEF), Deutschland

Cremers, Armin | B-IT Emeritus Research Group, Universität Bonn, Deutschland

Archer, Margaret | PASS-Akademiker, Professor, University of Warwick, UK

Reichberg, Greg | PASS-Akademiker, Professor, Peace Research Institute Oslo (PRIO), Norwegen

Sánchez Sorondo, Marcelo | Bischofskanzler der Päpstlichen Akademien der Wissenschaften (PAS) und Sozialwissenschaften (PASS)

### **Konferenzteilnehmer**

Akiyama, Nobumasa | Professor an der Graduate School of Law und der Graduate School of International and Public Policy, Hitotsubashi University, Japan

Battro, Antonio | PAS-Akademiker, Miembro de la Academia Nacional de Educación, Argentina

Clodic, Aurélie | Laboratoire d'Analyse et d'Architecture des Systèmes (LAAS) CNRS, Toulouse, Frankreich

De Backer, Koenraad | Abteilungsleiter, Direktion für Wissenschaft und Technologie und Innovationen, OECD, Paris, Frankreich

Dehaene, Stanislas | PAS-Akademiker, Professor, INSERM-CEA Cognitive Neuroimaging Unit, Collège de France, Orsay, Frankreich

Devillers, Laurence | Professor, Künstliche Intelligenz und Ethik, Paris-Sorbonne, Frankreich

Donati, Pierpaolo | PASS-Akademiker, PAM Professor für Soziologie, Universität Bologna, Italien

Dowek, Gilles | Forscher am INRIA, Professor an der ENS Paris-Saclay, Frankreich

Ema, Arisa | Assistant Professor, University of Tokyo; Gastwissenschaftlerin am RIKEN Center for Advanced Intelligence Project, Japan

Fischer, Sophie-Charlotte | Doktorandin, Zentrum für Sicherheitsstudien, ETH Zürich, Schweiz

Gabriel, Markus | Professor, Philosophisches Seminar, Universität Bonn, Deutschland

Hirzinger, Gerhard | Professor em., Technische Universität München, Deutschland

Kane, Angela | Senior Fellow des Wiener Zentrums für Abrüstung und Nichtverbreitung, Österreich

Léna, Pierre | PAS-Akademiker, Professor em., Université Paris VII Denis Diderot, Frankreich

Pasquale, Frank | Professor für Recht, University of Maryland, USA

Peylo, Christoph | Global Head of Bosch Center for Artificial Intelligence, mit Standorten in Sunnyvale, Bangalore und Renningen, Deutschland.

Quintarelli, Stefano | Vorsitzender des Lenkungsausschusses der italienischen Digitalagentur, Italien

Schröder, Wolfgang M. | Professor, Philosophie, Katholisch-Theologische Fakultät. Universität Würzburg, Deutschland

Singer, Wolf | PAS-Akademiker, Max-Planck-Institut für Hirnforschung, Frankfurt, Deutschland

Stachniss, Cyrill | Professor, Labor für Photogrammetrie und Robotik, Universität Bonn, Deutschland

Swett, Bruce A. | Professor, Johns Hopkins University Applied Physics Laboratory, Baltimore, USA, USA

Syse, Henrik | Professor am Peace Research Institute, Oslo, Norwegen

Torero, Maximo | Assistant Director General, Food and Agriculture Organization (FAO) der Vereinten Nationen, Rom, Italien

van Wynsberghe, Aimee | Professor, Ethik und Technologie, TU Delft, Niederlande

Zamagni, Stefano | Präsident der Päpstlichen Akademie der Sozialwissenschaften (PASS); Professor für Wirtschaftswissenschaft an der Universität Bologna, Italien