

Rome Call for AI Ethics * Römischer Aufruf zu einer KI-Ethik (Rome / Rom, 28.2.2020)

INTRODUCTION

“Artificial intelligence” (AI) is bringing about profound changes in the lives of human beings, and it will continue to do so. AI offers enormous potential when it comes to improving social coexistence and personal well-being, augmenting human capabilities and enabling or facilitating many tasks that can be carried out more efficiently and effectively. However, these results are by no means guaranteed. The transformations currently underway are not just quantitative. Above all, they are qualitative, because they affect the way these tasks are carried out and the way in which we perceive reality and human nature itself, so much so that they can influence our mental and interpersonal habits. New technology must be researched and produced in accordance with criteria that ensure it truly serves the entire “human family” (Preamble, Univ. Dec. Human Rights), respecting the inherent dignity of each of its members and all natural environments, and taking into account the needs of those who are most vulnerable. The aim is not only to ensure that no one is excluded, but also to expand those areas of freedom that could be threatened by algorithmic conditioning.

Given the innovative and complex nature of the questions posed by digital transformation, it is essential for all the stakeholders involved to work together and for all the needs affected by AI to be represented. This Call is a step forward with a view to growing with a common understanding and searching for a language and solutions we can share. Based on this, we can acknowledge and accept responsibilities that take into account the entire process of technological innovation, from design through to distribution and use, encouraging real commitment in a range of practical scenarios. In the long term, the values and principles that we are able to instil in AI will help to establish a framework that regulates and acts as a point of reference for digital ethics, guiding our actions and promoting the use of technology to benefit humanity and the environment.

Now more than ever, we must guarantee an outlook in which AI is developed with a focus not on technology, but rather for the good of humanity and of the environment, of our common and shared home and of its human inhabitants, who are inextricably connected. In other words, a vision in which human beings and nature are at the heart of how digital innovation is developed, supported rather than gradually replaced by technologies that behave like rational actors but are in no way human. It is time to begin preparing for more technological future in which machines will have a more important role in the lives of human beings, but also a future in which it is clear that technological progress affirms the brilliance of the human race and remains dependent on its ethical integrity.

EINLEITUNG

Die "Künstliche Intelligenz" (KI) bewirkt tiefgreifende Veränderungen im Leben der Menschen, und sie wird dies auch weiterhin tun. KI bietet ein enormes Potenzial, wenn es darum geht, das soziale Zusammenleben und das persönliche Wohlbefinden zu verbessern, die menschlichen Fähigkeiten zu erweitern und viele Aufgaben zu ermöglichen oder zu erleichtern, die wirksamer und erfolgreicher ausgeführt werden können. Diese Ergebnisse sind jedoch keineswegs garantiert. Die derzeit stattfindenden Transformationen sind nicht nur quantitativ. Sie sind vor allem qualitativ, denn sie beeinflussen die Art und Weise, wie diese Aufgaben ausgeführt werden und wie wir die Realität und die menschliche Natur selbst wahrnehmen, so dass sogar sie unsere geistigen und zwischenmenschlichen Gewohnheiten beeinflussen können. Neue Technologien müssen nach Kriterien erforscht und produziert werden, die sicherstellen, dass sie wirklich der gesamten "Menschheitsfamilie" dienen (Präambel zur Allgemeinen Erklärung der Menschenrechte), und dabei sind die jedem ihrer Mitglieder und jeder natürlichen Umwelt innewohnende Würde zu respektieren und die Bedürfnisse derer zu berücksichtigen, die am verletzlichsten sind. Es gilt nicht nur sicherzustellen, dass niemand ausgeschlossen wird, sondern auch jene Bereiche der Freiheit zu erweitern, die durch algorithmische Konditionierung bedroht sein könnten.

Angesichts des innovativen und komplexen Charakters der Fragen, die die digitale Transformation aufwirft, müssen unerlässlich alle beteiligten Akteure zusammenarbeiten und alle von der KI betroffenen Bedürfnisse vertreten sein. Dieser Aufruf ist ein Schritt vorwärts im Blick auf das Wachstum in einem gemeinsamen Verständnis und auf der Suche nach einer Sprache und nach Lösungen, die wir teilen können. Auf dieser Grundlage können wir Verantwortlichkeiten anerkennen und übernehmen, die den gesamten Prozess der technologischen Innovation berücksichtigen, vom Design bis hin zur Verbreitung und Nutzung, um so ein echtes Engagement in einer Reihe von praktischen Szenarien zu fördern. Langfristig werden die Werte und Prinzipien, die wir in die KI einbringen können, zur Schaffung eines Rahmens beitragen, der die digitale Ethik reguliert und als Bezugspunkt für sie dient, der unser Handeln lenkt und den Einsatz der Technologie zum Nutzen von Mensch und Umwelt fördert.

Wir müssen heute mehr denn je eine Perspektive garantieren, in der bei der Entwicklung der KI nicht die Technologie im Mittelpunkt steht, sondern eher das Wohl der Menschheit und der Umwelt, unseres gemeinsam geteilten Zuhause und seiner untrennbar miteinander verbundenen menschlichen Bewohner. Mit anderen Worten, eine Vision, in der Mensch und Natur das Herz für die Weise der Entwicklung digitaler Innovation bilden, unterstützt und nicht allmählich ersetzt durch Technologien, die sich wie rationale Akteure verhalten, jedoch in keiner Weise menschlich sind. Es ist an der Zeit, sich auf eine stärker technologische Zukunft vorzubereiten, in der Maschinen eine wichtigere Rolle im Leben der Menschen spielen werden, aber auch eine Zukunft, in der klar ist, dass der technologische Fortschritt die Leuchtkraft der menschlichen Rasse bestätigt und an deren ethische Integrität gebunden bleibt.

ETHICS

All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of fellowship (cf. Art. 1, Univ. Dec. Human Rights). This fundamental condition of freedom and dignity must also be protected and guaranteed when producing and using AI systems. This must be done by safeguarding the rights and the freedom of individuals so that they are not discriminated against by algorithms due to their "race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status" (Art. 2, Univ. Dec. Human Rights).

AI systems must be conceived, designed and implemented to serve and protect human beings and the environment in which they live. This fundamental outlook must translate into a commitment to create living conditions (both social and personal) that allow both groups and individual members to strive to fully express themselves where possible.

In order for technological advancement to align with true progress for the human race and respect for the planet, it must meet three requirements. It must include every human being, discriminating against no one; it must have the good of humankind and the good of every human being at its heart; finally, it must be mindful of the complex reality of our ecosystem and be characterised by the way in which it cares for and protects the planet (our "common and shared home") with a highly sustainable approach, which also includes the use of artificial intelligence in ensuring sustainable food systems in the future. Furthermore, each person must be aware when he or she is interacting with a machine.

AI-based technology must never be used to exploit people in any way, especially those who are most vulnerable. Instead, it must be used to help people develop their abilities (empowerment/enablement) and to support the planet.

EDUCATION

Transforming the world through the innovation of AI means undertaking to build a future for and with younger generations. This undertaking must be reflected in a commitment to education, developing specific curricula that span different disciplines in the humanities, science and technology, and taking responsibility for educating younger generations. This commitment means working to improve the quality of education that young people receive; this must be delivered via methods that are accessible to all, that do not discriminate and that can offer equality of opportunity and treatment. Universal access to education must be achieved through principles of solidarity and fairness.

Access to lifelong learning must be guaranteed also for the elderly, who must be offered the opportunity to

ETHIK

Alle Menschen sind frei und gleich an Würde und Rechten geboren. Sie sind mit Vernunft und Gewissen begabt und sollen einander im Geiste der Gemeinschaft begegnen (vgl. Art. 1, Allgemeine Erklärung der Menschenrechte). Diese Grundbedingung von Freiheit und Würde muss auch bei der Erzeugung und Nutzung von KI-Systemen geschützt und garantiert werden. Dies muss geschehen, indem Rechte und Freiheit des Einzelnen erhalten bleiben, so dass diese nicht diskriminiert werden durch Algorithmen aufgrund von "Rasse, Hautfarbe, Geschlecht, Sprache, Religion, politischen oder sonstigen Überzeugungen, nationaler oder sozialer Herkunft, Vermögen, von Geburt oder einem sonstigen Status" (Art. 2, Allgemeine Erklärung der Menschenrechte).

KI-Systeme müssen so konzipiert, gestaltet und eingesetzt werden, dass sie dem Menschen und der Umwelt, in der er lebt, dienen und diese schützen. Diese grundlegende Sichtweise muss sich in der Verpflichtung niederschlagen, Lebensbedingungen (soziale wie auch persönliche) zu schaffen, die es sowohl Gruppen als auch einzelnen Mitgliedern ermöglichen, sich nach Möglichkeit voll auszudrücken.

Damit die technologische Verbesserung mit echtem Fortschritt für die menschliche Rasse und der Achtung des Planeten einhergeht, muss sie drei Anforderungen erfüllen. Sie muss jeden Menschen einschließen und niemanden diskriminieren; sie muss das Wohl der Menschheit und das Wohl jedes Menschen als Herzmitte haben; schließlich muss sie die komplexe Realität unseres Ökosystems berücksichtigen und sich durch die Weise auszeichnen, wie sie den Planeten (unser "gemeinsam geteiltes Zuhause") mit einem hochgradig nachhaltigen Ansatz pflegt und schützt und dabei auch den Einsatz künstlicher Intelligenz einbezieht, um nachhaltige Ernährungssysteme in der Zukunft zu gewährleisten. Darüber hinaus muss jeder Mensch, wenn er mit einer Maschine interagiert, sich dessen bewusst sein.

KI-basierte Technologie darf niemals dazu benutzt werden, Menschen in irgendeiner Weise auszubeuten, insbesondere nicht diejenigen, die am verletzlichsten sind. Vielmehr muss sie eingesetzt werden, um Menschen bei der Entwicklung ihrer Fähigkeiten zu helfen (Empowerment) und den Planeten zu unterstützen.

ERZIEHUNG

Die Welt durch die Innovation der KI zu transformieren, bedeutet die Verpflichtung, eine Zukunft für die jüngeren Generationen und mit ihnen aufzubauen. Diese Verpflichtung muss sich niederschlagen in einem Engagement für die Bildung, die Entwicklung spezifischer Lehrpläne unter Einbeziehung verschiedener Disziplinen der Geisteswissenschaften, der Naturwissenschaften und der Technologie und in der Übernahme von Verantwortung für die Erziehung der jüngeren Generationen. Diese Verpflichtung bedeutet Arbeit, um die Qualität der Bildung, die junge Menschen erhalten, zu verbessern; dies muss mit Methoden geschehen, die für alle zugänglich sind, die nicht diskriminieren und die Chancengleichheit und Gleichbehandlung bieten können. Der universale Zugang zur Bildung muss durch Prinzipien der Solidarität und Fairness erzielt werden.

Der Zugang zu lebenslangem Lernen muss auch für ältere Menschen gewährleistet sein, denen die Möglichkeit geboten

access offline services during the digital and technological transition. Moreover, these technologies can prove enormously useful in helping people with disabilities to learn and become more independent: inclusive education therefore also means using AI to support and integrate each and every person, offering help and opportunities for social participation (e.g. remote working for those with limited mobility, technological support for those with cognitive disabilities, etc.).

The impact of the transformations brought about by AI in society, work and education has made it essential to overhaul school curricula in order to make the educational motto "no one left behind" a reality. In the education sector, reforms are needed in order to establish high and objective standards that can improve individual results. These standards should not be limited to the development of digital skills but should focus instead on making sure that each person can fully express their capabilities and on working for the good of the community, even when there is no personal benefit to be gained from this.

As we design and plan for the society of tomorrow, the use of AI must follow forms of action that are socially oriented, creative, connective, productive, responsible, and capable of having a positive impact on the personal and social life of younger generations. The social and ethical impact of AI must be also at the core of educational activities of AI.

The main aim of this education must be to raise awareness of the opportunities and also the possible critical issues posed by AI from the perspective of social inclusion and individual respect.

RIGHTS

The development of AI in the service of humankind and the planet must be reflected in regulations and principles that protect people – particularly the weak and the underprivileged – and natural environments. The ethical commitment of all the stakeholders involved is a crucial starting point; to make this future a reality, values, principles, and in some cases, legal regulations, are absolutely indispensable in order to support, structure and guide this process.

To develop and implement AI systems that benefit humanity and the planet while acting as tools to build and maintain international peace, the development of AI must go hand in hand with robust digital security measures.

In order for AI to act as a tool for the good of humanity and the planet, we must put the topic of protecting human rights in the digital era at the heart of public debate. The time has come to question whether new forms of automation and algorithmic activity necessitate the development of stronger responsibilities. In particular, it will be essential to consider some form of

werden muss, während des digitalen und technologischen Übergangs auf Offline-Dienste zuzugreifen. Darüber hinaus können diese Technologien sich als enorm nützlich erweisen, um Menschen mit Behinderungen beim Lernen und bei der Erlangung von mehr Unabhängigkeit zu helfen: Inklusive Bildung bedeutet daher auch, dass KI zur Unterstützung und Integration jeder einzelnen Person eingesetzt wird, indem sie Hilfe und Möglichkeiten zur sozialen Teilhabe bietet (z.B. Fernarbeit für Menschen mit eingeschränkter Mobilität, technologische Unterstützung für Menschen mit kognitiven Behinderungen usw.).

Die Auswirkungen der durch die KI hervorgerufenen Transformationen in Gesellschaft, Arbeit und Bildung haben es notwendig gemacht, die Lehrpläne der Schulen zu überarbeiten, um das pädagogische Motto "niemand bleibt zurück" zu verwirklichen. Im Bildungssektor sind Reformen notwendig, um hohe und objektive Standards zu etablieren, die individuelle Ergebnisse verbessern zu können. Diese Standards sollten sich nicht auf die Entwicklung von digitalen Fähigkeiten beschränken, sondern sich stattdessen darauf konzentrieren, dass jede Person ihre Fähigkeiten voll zum Ausdruck bringen kann, sowie darauf, dass der Einsatz für das Wohl der Gemeinschaft erfolgt, auch wenn daraus kein persönlicher Nutzen gezogen werden kann.

Wenn wir die Gesellschaft von morgen entwerfen und planen, muss der Einsatz von KI nach Formen des Handelns erfolgen, die sozial orientiert, kreativ, verbindend, produktiv, verantwortungsvoll und in der Lage sind, das persönliche und soziale Leben der jüngeren Generationen positiv zu beeinflussen. Die sozialen und ethischen Auswirkungen der KI müssen auch im Mittelpunkt der Bildungsaktivitäten zur KI stehen.

Das Hauptziel dieser Bildung muss es sein, das Bewusstsein für die Chancen und auch für die möglichen kritischen Fragen zu wecken, die die KI aus der Perspektive der sozialen Integration und des individuellen Respekts aufwirft.

RECHTE

Die Entwicklung der künstlichen Intelligenz im Dienste der Menschheit und des Planeten muss sich in Vorschriften und Prinzipien widerspiegeln, die die Menschen – insbesondere die Schwachen und Unterprivilegierten – und die natürliche Umwelt schützen. Das ethische Engagement aller beteiligten Akteure ist ein entscheidender Ausgangspunkt; um diese Zukunft Wirklichkeit werden zu lassen, sind Werte, Prinzipien und in einigen Fällen auch gesetzliche Regelungen absolut unerlässlich, um diesen Prozess zu unterstützen, zu strukturieren und zu lenken.

Um KI-Systeme zu entwickeln und einzusetzen, die der Menschheit und dem Planeten zugute kommen und gleichzeitig als Mittel für den Aufbau und die Erhaltung des internationalen Friedens fungieren, muss die Entwicklung der KI mit robusten digitalen Sicherheitsmaßnahmen einhergehen.

Damit die KI als Mittel zum Wohle der Menschheit und des Planeten wirken kann, müssen wir das Thema des Schutzes der Menschenrechte im digitalen Zeitalter in den Mittelpunkt der öffentlichen Debatte stellen. Die Zeit ist gekommen, um zu hinterfragen, ob neue Formen von Automatisierung und algorithmischen Aktivitäten nicht die Entwicklung stärkerer Verantwortlichkeiten notwendig machen. Insbesondere wird

"duty of explanation": we must think about making not only the decision-making criteria of AI-based algorithmic agents understandable, but also their purpose and objectives. These devices must be able to offer individuals information on the logic behind the algorithms used to make decisions. This will increase transparency, traceability and responsibility, making the computer-aided decision-making process more valid.

New forms of regulation must be encouraged to promote transparency and compliance with ethical principles, especially for advanced technologies that have a higher risk of impacting human rights, such as facial recognition.

To achieve these objectives, we must set out from the very beginning of each algorithm's development with an "algor-ethical" vision, i.e. an approach of ethics by design. Designing and planning AI systems that we can trust involves seeking a consensus among political decision-makers, UN system agencies and other intergovernmental organisations, researchers, the world of academia and representatives of non-governmental organizations regarding the ethical principles that should be built into these technologies. For this reason, the sponsors of the call express their desire to work together, in this context and at a national and international level, to promote "algor-ethics", namely the ethical use of AI as defined by the following principles:

1. **Transparency:** in principle, AI systems must be explainable;
2. **Inclusion:** the needs of all human beings must be taken into consideration so that everyone can benefit and all individuals can be offered the best possible conditions to express themselves and develop;
3. **Responsibility:** those who design and deploy the use of AI must proceed with responsibility and transparency;
4. **Impartiality:** do not create or act according to bias, thus safeguarding fairness and human dignity;
5. **Reliability:** AI systems must be able to work reliably;
6. **Security and privacy:** AI systems must work securely and respect the privacy of users.

These principles are fundamental elements of good innovation.

Rome, February 28th, 2020

Quelle: <http://www.academyforlife.va/content/pav/it/events/workshop-intelligenza-artificiale.html>

Die Päpstliche Akademie für das Leben, Microsoft, IBM, die FAO und die italienische Regierung unterzeichneten den Aufruf für eine KI-Ethik, ein Dokument, das geschaffen wurde, um einen ethischen Ansatz zur künstlichen Intelligenz zu unterstützen und das Gefühl der gemeinsamen Verantwortung von Organisationen, Regierungen und Institutionen zu fördern, mit dem Ziel, eine Zukunft zu sichern, in der digitale Innovation und technologischer Fortschritt dem Menschen und der Kreativität dienen und nicht deren allmählicher Ersetzung.

Erstunterzeichner: Msgr. Vincenzo Paglia, Präsident der Päpstlichen Akademie auf Lebenszeit; Dr. Brad Smith, Präsident von Microsoft; Dr. John Kelly III, Executive Vice President IBM; Dr. Dongyu Qu, Generaldirektor der FAO, Minister Paola Pisano für die italienische Regierung.

es wesentlich sein, eine Form der "Erklärungspflicht" in Betracht zu ziehen: Wir müssen darüber nachdenken, nicht nur die Entscheidungskriterien der KI-basierten algorithmischen Akteure verständlich zu machen, sondern auch ihren Zweck und ihre Ziele. Diese Akteure müssen in der Lage sein, den Individuen Informationen über die Logik hinter den zur Entscheidungsfindung eingesetzten Algorithmen zu liefern. Dies wird zu mehr Transparenz, Nachvollziehbarkeit und Verantwortung beitragen und dem computergestützten Entscheidungsprozess mehr Geltung verschaffen.

Neue Formen der Regulierung müssen angeregt werden, um Transparenz und Einhaltung ethischer Grundsätze zu fördern, insbesondere bei hochentwickelten Technologien, die ein höheres Risiko der Auswirkung auf die Menschenrechte bergen, z.B. die Gesichtserkennung.

Um diese Ziele zu erreichen, müssen wir von Anfang an bei der Entwicklung jedes Algorithmus eine "algorithische" Vision verfolgen, d.h. einen Ansatz der "Ethik durch Design". Das Entwerfen und Planen von KI-Systemen, denen wir vertrauen können, beinhaltet die Suche nach einem Konsens zwischen den politischen Entscheidungsträgern, den Organen der Vereinten Nationen und anderen zwischenstaatlichen Organisationen, Forschern, der Welt der Wissenschaft und Vertretern von Nichtregierungsorganisationen hinsichtlich der ethischen Prinzipien, die in diese Technologien eingefügt werden sollten. Aus diesem Grund bringen die Befürworter des Aufrufs ihren Wunsch zum Ausdruck, in diesem Zusammenhang und auf nationaler und internationaler Ebene zusammenzuarbeiten, um eine "Algor-Ethik" zu fördern, d.h. die ethische Nutzung der KI, wie sie durch folgende Prinzipien definiert wird:

1. **Transparenz:** KI-Systeme müssen grundsätzlich erklärbar sein;
2. **Inklusion:** Die Bedürfnisse aller Menschen müssen berücksichtigt werden, damit alle davon profitieren und allen Menschen die bestmöglichen Bedingungen geboten werden können, um sich auszudrücken und sich zu entwickeln;
3. **Verantwortung:** Diejenigen, die den Einsatz von KI konzipieren und einsetzen, müssen mit Verantwortung und Transparenz vorgehen;
4. **Unvoreingenommenheit:** Es gilt nicht parteiisch zu schaffen oder zu handeln, um Fairness und Menschenwürde zu gewährleisten;
5. **Verlässlichkeit:** KI-Systeme müssen zuverlässig zu arbeiten vermögen;
6. **Sicherheit und Privatsphäre:** KI-Systeme müssen sicher arbeiten und die Privatsphäre der Benutzer respektieren.

Diese Prinzipien sind grundlegende Elemente einer guten Innovation.

Rom, 28. Februar 2020