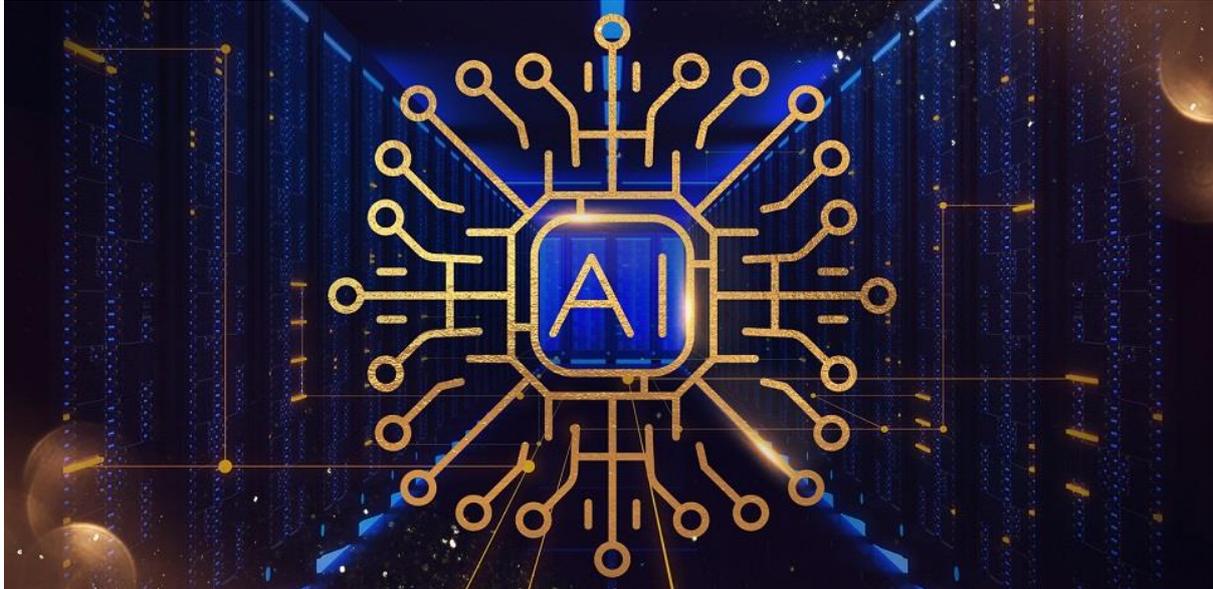


# OXFORD POLITICAL REVIEW

## [Artificial Intelligence: Implications for human dignity and governance](#)



[Nayef Al-Rodhan](#) March 27, 2021

### **Nayef Al-Rodhan**

Recent years have seen a surge in discussions about the impacts of artificial intelligence (AI). These debates have predominantly featured issues related to autonomy in driverless cars, or the moral dilemmas of deploying ‘killer robots’, though the reach and impact of AI-based technologies is, of course, far more widespread. AI is a widely common feature of our daily lives, present in systems that monitor our online searches and spam us with [advertising](#), impact voters’ decisions, but also in medicine, and in algorithms which determine [police profiling or foreclosing on a mortgage](#). AI is also at the forefront of the ongoing [Fourth Industrial Revolution](#), and it is estimated its contribution to the global economy could reach [\\$13 trillion by 2030](#).

The hype around AI has gone as far as to [compare its transformative potential](#) to that of electricity over the past century. There is then the security angle. The geopolitical implications, particularly in the context of the US-China rivalry – which are currently lead players in AI – will weigh heavily on the future of international relations. However, understanding AI’s impacts goes beyond a strictly geopolitical lens.

AI is also poised to impact statecraft in profound ways, altering or reshaping state policies and governance. I previously framed this understanding of state power in the 21<sup>st</sup> century as *meta-geopolitics*, which takes a more holistic view at state power as a combination of seven capacities: *social & health issues, domestic politics, economics, environment, science & human potential, military & security issues, and international diplomacy*. Together, these interrelated domains contribute to, and shape, national power. [Developments in AI](#) are set to impact each of these domains, bringing unique disruptions, and opportunities, in every aspect of modern statecraft. AI is becoming critical as an enabler of power projection within these sectors, as well as a potential threat

to them. Furthermore, artificial intelligence will inevitably transform the relationship between states and citizens, and impact human dignity in profound ways. The safe and sustainable use of AI going forward can only be achieved when the risks for human dignity are mitigated.

In the following section, I unpack the multifaceted implications of AI on the meaning of power, governance, and on human dignity. Importantly, I do not flag only the risks associated with the use of AI, but also areas of opportunities. A balanced perspective highlights the complex uses and ramifications of AI-based technologies, and the fine line between risks and benefits. I conclude by offering some concrete ideas for governance.

### **AI And Dignity-Based Governance**

As [emotional, amoral, and egoistic](#) beings (a neuro-philosophical account of human nature I developed previously), human dignity is the most fundamental of human needs. Dignity, even more so than freedom (dignity encompasses that and more), is critical for sustainable governance, and in a basic sense, essential to sustaining our social existence. I define dignity holistically to mean much more than the absence of humiliation but also the presence of recognition. It is a comprehensive set of [nine dignity needs](#), which are *reason, security, human rights, transparency, justice, accountability, opportunity, innovation, and inclusiveness*. AI impacts these needs in complex ways, by reinforcing or endangering them (and sometimes, both).

For **reason**, which refers to the absence of dogma (especially relevant in regimes that claim an absolute monopoly on truth), AI can have a positive impact by introducing rational techniques into decision-making processes. At the same time, the use of [big data](#) in policy-making may not always produce an outcome backed by reason or impartiality, but rather one potentially vulnerable to emotions. Products of AI such as [deepfakes](#) could also blur the lines between true and false, further weakening trust. Finally, with the increased [merger of humans and machines](#), AI could take decisions on our behalf, or sway our decisions, thus challenging our belief in the value of human reason.

**Security** is another fundamental need for human beings. On the one hand, AI can provide better information and help in the development of [encryption systems](#). On the other, cyber-attacks, prone to become more frequent and target crucial infrastructures such as healthcare systems, will increase anxiety across societies, with far-reaching implications for public order.

In terms of opportunities for **human rights**, AI systems could help in the prevention, detection and monitoring of violations, for instance by analyzing [satellite imagery and social media](#) content. On the other hand, fundamental freedoms, such as the right to privacy, will be threatened by large scale data collection and new methods of surveillance and policing. In fact, governments may use AI to monitor social media activity, as well as to trace and identify people through [facial recognition](#).

The human need for **transparency** implies that authorities and private companies must provide clear information on their activities to remain legitimate actors. The collection of data to feed AI systems, without clear indications on its intended or potential use, endanger transparency, and even more so in the context of the rise of [tech giants](#) which develop such technologies outside the eyes of the public or governments regulations.

**Justice** is fundamental to sustainable governance and the sense of fairness is deeply relevant to human nature (I have elaborated on its neurochemical representation in [another article](#)). In concrete terms, AI could help judicial institutions through investigations techniques such as [DNA analysis](#). The use of AI technologies has, however, also been found to produce discriminatory outcomes in

the [justice system](#) due to AI's vulnerability to [biases](#), and technologies such as facial recognition have been known to identify specific ethnicities in a discriminatory manner.

**Accountability** is essential for consolidating trust and security in society. In this regard, AI may assist in identifying authors of malicious acts, but this comes, too, with serious accountability challenges. The overreliance on such systems may lose sight of the growing errors and biases the systems will learn and replicate in time (for instance, an algorithm in a [recruitment system developed at Amazon](#) to identify the world's best engineers had learnt to exclude women from its scoring process). Furthermore, accountability is even more complicated when [incidents](#) occur, making the human chain of responsibility hardly traceable.

A feeling of equal **access to opportunities** is necessary to ensure social cohesion. The shift towards more digital societies will inevitably create inequalities, as it risks reducing [employment](#) in some sectors while creating new jobs elsewhere. More importantly, [human enhancement](#) technologies, some of which will integrate AI systems, are poised to impact the future of work, with implications across sectors and professions. One concern is that these enhancement technologies will not be distributed fairly, thus giving rise to a society not based on merit. At the same time, as a Royal Society [report](#) highlights, it can also bring benefits by enabling some professionals to resume work or work in easier conditions. However, the public management and regulation of access to such technologies will be critical.

The development of AI certainly also fulfills our need for **innovation**, bringing new opportunities in several domains, from education and [healthcare](#) to the military. AI, on the whole, can accelerate innovations and R&D, and shape the future of other technologies. What remains as a challenge, however, is wide access and transfer of knowledge and technology across the world, and particularly in emerging economies.

There are several ways in which AI could harm our need for **inclusiveness** and become divisive. This happens in everyday situations, as AI tools target individuals online with [tailored content](#), thus reducing exposure to different points of view and reinforcing biases. In many other instances, [biases](#) embedded in algorithms continue to lead to discriminatory practices and social injustice. In the medium to long run, different allocations of resources and shifts in the job market in the context of automation are bound to sharpen inequalities. In even more extreme forms, the advent of [human enhancement technologies](#) will trigger divisions at the societal level and risk split citizens into "in-groups" and "out-groups".

## **Way Forward & Policy Recommendations**

This brief survey of risks and opportunities of AI reminds us that only by promoting mechanisms of governance and strict oversight mechanisms can the positive features of AI surpass the risks associated with it. A consideration for human dignity must serve as the fundamental goal for regulating AI but that requires, in practical terms, a series of policy and legal commitments. Importantly, this gives a voice, and responsibility, to a wide array of actors.

### **For tech companies**

#### **1. Establish ethics committees**

Ethics committees must be set up by private actors to reflect on challenges brought by AI, and promote the adoption of ethical principles at all levels, from the early stages of [coding](#).

#### **2. Strengthen data protection systems**

Tech companies developing AI systems must commit to ensure higher security levels for data storage and use. This could involve increasing their security R&D budgets and supporting new projects on encryption techniques.

#### **For states**

##### **1. Bridge knowledge gaps and ensure trust between stakeholders and policymakers**

Public authorities must be proactive in bridging the knowledge gaps between policymakers, technical experts, constituencies and tech companies. While there are significant knowledge gaps at the national policy level, there is also an important need for regulation. This process could involve public consultations with actors of the sectors concerned, possibly through the creation of a cross-sectorial commission which could be responsible for compiling existing knowledge and standards on AI. In the process, citizens must also become more familiarized with AI through [training and education](#) programs.

##### **2. Adapt national policies**

A public authority could also be established, responsible for translating the information gathered from these consultation processes into policy guidelines for all the categories affected. Such consultations could also adopt scenario-making strategies, aim to predict developments over the next few years and help to allocate national budgets.

##### **3. Update national laws and regulations**

Since innovation is occurring at incredible speed, national laws and regulations will require updating. Expert commissions should be created to inform relevant stakeholders and governments should propose the review of relevant existing bodies of laws. Such legislative changes need to address, among others, liabilities for giant tech companies, whose expanding powers create disproportionate advantages and limit accountability. In the US, for example, [Section 230](#) of the Communications Decency Act passed in 1996 (in the early days of the internet) effectively offers a liability shield for internet companies, and continues to create controversy in light of these companies' outstanding reach. Going forward, guaranteeing the respect for constitutionally protected rights will be critical, and no private company should be exempt from liabilities.

##### **4. Invest in skill-adaptability programs**

AI will bring systemic changes to production chains and will require an adaptation of employees' skills. States must play central parts in investing in skill-adaptation programs for their citizens, which would reduce the risk of an increase in unemployment rates.

#### **For international institutions**

##### **1. Put AI at the forefront of international agendas**

International institutions must also work to influence agendas and ensure that AI regulation is put at the forefront. They could contribute to this process through the production of technical reports and engagement with public and private actors to ensure participation. More concretely, as suggested by Eleonore Pauwels, a [UN Global Foresight Observatory for AI Convergence](#) could be created to gather public and private actors to build scenarios, map stakeholders and develop approaches for innovation and prevention at the global level.

##### **2. Address inequalities introduced by AI**

These entities should ensure that AI development is also understood through the ethical and economic challenges it will bring to the world population, especially the rise in inequalities. Through reports and engagement with relevant actors, they can assist the international community to highlight the main challenges of AI, and particularly the impact of AI on vulnerable groups.

### **3. Develop adaption plans through in-house AI Task Forces**

Finally, to address the challenges brought by AI within their own bodies, in-house task forces could be created to design and manage adaptation plans. They could be formed by technical experts, human resources and management team members as well general staff.

The advent of artificial intelligence, and its accelerated growth in the past decade, is set to impact the future of humanity, from the workplace to the battlefield, and from local governance to global affairs. The attainment and respect of human dignity will be, and in some sectors already is, a critical area of concern for artificial intelligence, with potentially grave consequences for our freedoms, though – as I showed above – there are also opportunities that can be leveraged.

Nevertheless, against a backdrop of profound alteration to the meanings of power, states retain unique prerogatives to create regulatory frameworks for setting the course of AI developments. Even as they appear weakened in the face of private companies' immense power and grip in the global market, states are uniquely positioned to ensure that artificial intelligence and related technologies do not set us on a dangerous course of loss of dignity and disproportionate disempowerment against giant tech companies.

The complexities and challenges of regulating AI technologies surely require input and expertise from a wider array of actors, yet states' role in shaping a course for the future stands out. Its underlying scope must be the unwavering respect and guarantee of the nine human dignity needs outlined above, *for all, at all times and under all circumstances*. This is the ultimate pre-requisite for effective, accountable, equitable, and sustainable governance in our uncertain and intrusive future.

---

Professor Nayef Al-Rodhan, is a Philosopher, Neuroscientist and Geostrategist. He is *Honorary Fellow*, St. Antony's College, Oxford University, UK, *Head of the Geopolitics and Global Futures Program*, Geneva Center for Security Policy, Switzerland, *Senior Research Fellow*, Institute of Philosophy, School of Advanced Study, University of London, UK, and *Member of the Global Future Council on Frontier Risks* at the World Economic Forum. His research focuses on the interplay between: Analytic Neurophilosophy and Policy, History, Geopolitics, Global Futures, Outer Space Security, Global Trans-cultural Relations, Conflict, Global Security, Disruptive technologies, International Relations & Global Order.