

# Artificial Intelligence *ante portas*: Legal & ethical reflections

In view of the mushrooming number of artificial intelligence (AI) applications and the ability of AI to act autonomously and in unforeseeable ways, the growing interest in the ethical and legal aspects of AI is hardly surprising. AI appears to hold the potential to bring forward disruptive changes to the law as the existing legal structures struggle to cope with the increasing opacity and self-learning development of autonomous machines and the associated concerns about control and responsibility. Reshaping the current legal framework in accordance with the new technological realities will predominantly depend on the technological progress and commercial uptake of AI, as more powerful AI may even require a fundamental paradigm switch. Given the foreseeable pervasiveness of AI, it is both legitimate and necessary to pose the question about how this constellation of technologies should be defined, classified and translated in legal and ethical terms.

## Legal challenges/dimensions

As AI systems are used in more common and consequential contexts and will soon be applied in [safety-critical applications](#), such as clinical decision support and autonomous driving, there is increasing attention on whether, how, and to what extent they should be regulated. As a transformative technology that is characterised by high complexity, unpredictability and autonomy in its decision-making and learning capacities, AI has the potential to challenge traditional notions of legal personality, individual agency and responsibility. The introduction of self-running and self-enforcing AI systems that can operate independently from their creators or operators and are equipped with adaptive and learning abilities that allow them to learn from their own variable experience and interaction with their environment in a unique and unforeseeable manner, may signify a significant shift in the fundamental underpinnings of law. Not least given that the rule of law is traditionally based on predictability and the legal obligation to compensate because of an unlawful injury to a person or property.

Legal systems currently face an unprecedented range of AI-related challenges associated, among other things, with the need to prevent algorithmic bias, safeguard human control over the operation of automated intelligent systems and hold such intelligent systems accountable. However, is the law the right instrument to address bias or restore human presence in purely automated systems? Is bias a legal or a cultural and social matter? Is anticipative regulation justified for certain classes of risks and/or AI applications? Should legal systems resort to the precautionary principle to avoid dealing with the risks of a rapidly changing technological trajectory? Or should legal experimentation be enabled, such as the introduction of regulatory 'sandboxes' that could contain possible technological wrongdoings? Are technology-based initiatives necessary, or is a product-specific approach more appropriate? Can a line be drawn between legal and ethical guidance in this context?

## Defining AI

Defining the precise object of regulation in dynamic technological domains is a challenge in itself. Given that AI is still an open-ended notion that refers to a very wide range of products and applications,

there is no transnational agreement on a commonly accepted working definition, neither at the technical nor the legal/policy level. As there is no legal and political consensus over what AI is, a plurality of definitions has emerged in Europe and worldwide that are either too inclusive or too sector-specific. This fragmented conceptual landscape may prevent the immediate development of a *lex robotica* and possibly undermine all efforts to create a common legal nomenclature, which is particularly instrumental for the drafting, adoption and effective implementation of binding legal norms. Alternatively, a broad and technology-neutral definition that is based on the fulfilment of a variety of structural criteria, including the level of autonomy and the function, may be a more plausible option.

If no definitional scheme exists that can embrace all possible uses and applications of this technology, then the introduction of flexible instruments such as delegated acts, sunset clauses and experimental legislation may be necessary to address the [pacing problem](#) – adapting rules to changing technological circumstances and keeping pace with rapid technological developments. Retaining such flexibility is important to encouraging technological innovation in the field of AI, as well as to ensure that Union legislation is developed in a careful and inclusive manner.

The problem of definitional ambiguity is closely associated with the issue of AI's legal classification and the categorisation of its various applications. Should AI products and systems be approached under the umbrella of traditional legal categories, or are we simply experiencing the gradual creation of an entirely new domain of critical legal thinking that may trigger a shift from the traditional notion of code as law to a novel conceptualisation of law as code?

Finally, as autonomous AI applications develop, questions arise about whether they should be granted legal personhood. The development of the notion of 'robot personhood' for smart robots and the attribution of legal personality to AI in its smart/strong sense may appear beneficial, as it could restore the chain of causation and limit the liability of the owner. At the same time, the creation of such a legal fiction does not seem to meet the traditional criteria of personhood and is based on [what many experts consider](#) to be 'an overvaluation of the actual capabilities of even the most advanced robots, a superficial understanding of unpredictability and self-learning capacities, a robot perception distorted by science fiction and a few recent sensational press announcements'.

## Accountability

The complexity of AI models combined with the use of various forms of automation and algorithmic bias raises challenges for the transparency and accountability of machine learning and AI systems. Thus, a very important legal concern relates to the need for algorithmic fairness, transparency and accountability in the context of AI. Shedding light on the course of action followed for the collection, use and processing of personal data can be hindered by the need to protect those commercial actors that develop the algorithms and data processing mechanisms on the basis of the need to protect trade secret rights.

Transparency, in terms of disclosing the algorithmic code, does not however safeguard whether, and under which conditions, the algorithm was actually used in the respective decision-making system, or whether it 'behaved' as it was initially programmed. Thus, explainability may be better positioned to facilitate the examination of the actual terms of interface between the inputs and outputs of algorithms, and the underlying assumptions, as well as how datasets were trained and implemented. The deployment of a 'right to explanation' of algorithmic decisions would entitle users to receive clarification of the reasoning as to how a decision concerning them was reached with the aid of AI; to understand the value of the data produced in the frame of an AI-driven algorithmic decision-making system; and to become properly informed of the storage and destruction procedures, as well as the terms, of the required informed consent process.

Implementing such a right may shed light on the operation of AI systems, along with auditing of the source code, and introducing strong requirements for transparency and accountability should complement the above-mentioned efforts. The requirement for data controllers to provide data

subjects with 'meaningful information about the logic involved' in an automated decision-making process was recently introduced by the [General Data Protection Regulation](#) (GDPR). Uncovering the reasoning and main assumptions behind AI-based decision-making is a major legal concern that transcends several domains of the digital economy. It must always be possible to reduce the [AI system's](#) computations to a form comprehensible by humans, and to equip advanced robots with a 'black box' which could record data on every transaction carried out by the machine, including the logic that contributed to its decisions.

In other words, particular legal attention should be paid not only to the need to balance between the requirements of the GDPR and the [Trade Secrets Directive](#), but also to render algorithmic literacy an essential part of any regulatory attempt to control algorithmic decision-making that may lead and/or entail social discrimination or undermine the protection of fundamental human rights. Improving the accessibility and readability of the procedure of data processing will also pave the way for a better understanding of the type of data being processed, to whom it belongs, how the data spread, and the purpose for which the data are analysed.

## Liability

The question of how to ensure AI systems are transparent and accountable in their operations is closely related to the debate about the allocation of liability in the context of AI. The autonomous and unpredictable character of the operation of AI systems could lead to questions about causation, unforeseen circumstances, and fuzzy attribution of liabilities. Who will be held liable when an AI application causes physical or moral harm? Who will be held responsible if an AI system proposes a plan that proves damaging, the manufacturer or those actors that end up implementing algorithms? However, is it possible for AI developers to uphold and protect the right to privacy and obtain clear, unambiguous and informed consent, especially given the placement of some of its applications in traditionally protected and private spheres? Can an algorithm be sued for malpractice?

Due to the rapid development of certain autonomous and cognitive features, civil law will have to treat this technology differently, and possibly depart from its traditional liability theories, including product liability, negligence and strict liability models. A need therefore exists to develop a proportionate civil liability regime that could ensure a clear division of responsibilities among designers, manufacturers, service providers and end users. The European Parliament resolution of 16 February 2017 with recommendations to the Commission on 'Civil law rules on robotics' proposed, among other things, the establishment of a compulsory insurance scheme, of a compensation fund and of a Union register, and suggested the creation of a specific legal status for robots in the long-term, so that at least the most sophisticated autonomous robots could be established as having the [status of electronic persons](#) responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently.

## Data protection/privacy

AI requires access to extensive sets of data, in many cases, sensitive or protected data including data on race, ethnicity, gender and other sensitive attributes. Legal concerns arise regarding the extrapolative power of AI, the volume of data required to effectively develop algorithms and machine learning patterns, as well as with regard to the market concentration tendencies in the AI field. AI's ability to analyse data and identify users may in fact increase the sensitivity of data that was previously considered sufficiently anonymous. Despite the gradual development of anonymisation techniques that enable privacy-preserving big data analysis, AI poses a growing threat to the right of human beings to form their own opinions and take autonomous decisions. Particular attention should be paid to AI's capacity to use personal and non-personal data to [sort and micro-target people](#), to identify individual vulnerabilities and exploit accurate predictive knowledge. Any initiative that encourages open and free flow of data should take account of the legal need to comply with the principles of data minimisation, the right to obtain an explanation of a decision based on automated processing, the principles of

privacy by design and by default and the principles of proportionality, necessity, data minimisation, and purpose limitation.

## Safety/security

Autonomous vehicles and other AI-related systems may pave the way for security breaches, cyber-attacks or misuse of personal data, particularly when it involves the collection and processing of a large amount of data. Significant safety risks and vulnerabilities are associated with the possibility that AI applications that are integrated into the human body may be hacked as this may endanger human health, cognitive liberty or even mental integrity and human life.

Misuse of AI may not only threaten digital security and physical and public safety, but might also pose a risk to democracy in the form of disinformation campaigns and election hacking. [AI products](#) should be subject to product safety and consumer protection rules that ensure, where appropriate, minimum safety standards and address the risk of accidents resulting from interaction with, or working in proximity to, humans. In this respect, cybersecurity rules and legal safeguards are needed to guarantee that data is not maliciously corrupted or misused, which may undermine industry and consumer trust in AI.

## Socio-ethical challenges

Beyond the legal challenges, AI poses many ethical considerations. The use of AI for monitoring or even predicting human behaviour [risks](#) stigmatisation, reinforcing existing stereotypes, social and cultural segregation and exclusion, subverting individual choice and equal opportunities. AI's potential for empowerment, the risks associated with entering into 'filter bubbles' or with using social scoring methodologies through the use of AI and the affordability and accessibility of AI services are associated with concerns about human safety, health and security, freedom, privacy, integrity, dignity, self-determination and non-discrimination.

The increased use of AI-based algorithmic decision-making in the domains of financial services, banking and criminal justice without the involvement of human judgement or due process can reinforce harmful social stereotypes against particular minority groups and amplify racial and gender biases. This practise has been criticised by several EU-wide institutional actors, such as the [Council of Europe](#), the [European Data Protection Supervisor](#) and the [European Union Agency for Fundamental Rights](#).

The advent of strong AI, which penetrates legal categories, and triggers the reconsideration of traditional legal terms such as autonomy and privacy, raises questions about the capacity of EU law to strike the right balance between technology as a regulatory object or category and technology as a regulatory agenda-setter. Strong AI developments also raise the need to acknowledge that codes codify values and cannot be treated as a mere question of engineering. The ethical issues associated with the social power of algorithms accompany questions about the intergenerational digital divide and may affect the enjoyment of the right to life, the right to a fair trial, the right to privacy, freedom of expression and workers' rights. There are strong ethical, psychological and legal concerns about the autonomy of smart robots, and their impact on the doctor-patient relationship in healthcare applications, which have not yet been properly addressed at EU level, in particular as regards the protection of patients' personal data, liability, and the resulting new economic and employment relationships.

A strict and efficient guiding ethical framework for the development, design, production and use of algorithms is therefore needed to complement the existing national and Union *acquis*. The guiding ethical framework, which should safeguard human oversight over automated and algorithmic decision-making, should be based on the principles and values enshrined in the Charter of Fundamental Rights, such as human dignity, equality, justice and equity, non-discrimination, informed consent, private and family life and data protection, as well as on other underlying principles and values

of Union law, such as non-stigmatisation, transparency, autonomy, individual responsibility and social responsibility, and on existing ethical practices and codes.

## EU-wide initiatives

The rise of AI in Europe has so far occurred in a regulatory void. With the exception of the development of ethical codes of conduct and guidelines, very few legal initiatives have been taken that view AI in a holistic manner or at systemic level. No horizontal rule or judicial decision has been adopted that specifically addresses the unique challenges raised by AI.

On 10 April 2018, a [Declaration of Cooperation on AI](#) was signed by 25 countries, who agreed to 'work together on the most important issues raised by AI; from ensuring Europe's competitiveness in the research and deployment of AI, to dealing with social, economic, ethical and legal questions'. At EU level, the European Parliament [called](#) on the European Commission to assess the impact of artificial intelligence and made wide-ranging recommendations on civil law rules on robotics in February 2017. Furthermore, the Parliament adopted an own-initiative-report on a '[Comprehensive European industrial policy on artificial intelligence and robotics](#)' in February 2019. The European Economic and Social Committee issued an [opinion](#) on AI in May 2017. In the meantime, the European Council of October 2017 [invited](#) the Commission to put forward a European approach to AI.

The European Commission highlighted the importance of being in a leading position in the development of AI technologies, platforms and applications, in its [mid-term review of the digital single market strategy](#) published in May 2017. Finally, the European Commission adopted a [communication on 'Artificial intelligence for Europe'](#) on 25 May 2018, laying down the European approach to benefiting from the opportunities offered by AI and addressing the new challenges AI poses. The Commission proposed a three-pronged approach: increasing public and private investment; preparing for socio-economic changes brought about by AI; and ensuring an appropriate ethical and legal framework.

With regard to the ethical and legal framework, the Commission is expected to propose AI ethical guidelines, issue a guidance document on the interpretation of the [Product Liability Directive](#) by mid-2019, and to undertake studies and research and formulate policy responses to the challenges posed by AI regarding liability, safety, the internet of things (IoT), robotics, algorithmic awareness, consumer and data protection. In June 2018, the Commission appointed 52 experts to a new High Level Expert Group on Artificial Intelligence (AI HLEG), comprising representatives mostly from academia and industry, which aims at supporting the implementation of the European strategy on AI. In December 2018, the AI HLG published its [first draft of EU AI ethical guidelines](#). The draft guidelines were available and open for consultation until 1 February 2019. In December 2018, the Commission also published a [coordinated plan on artificial intelligence](#), aiming at coordinating the Member States' approaches, with respect to their national AI strategy objectives, their learning and skilling programmes, the financing mechanisms available but also in relation to the review of the existing legislation and ethical guidance. The Council of Ministers adopted [conclusions](#) in relation to the coordinated plan on the development and use of '[Artificial intelligence made in Europe](#)' in February 2019.

In terms of specific completed regulation on AI, the EU's regulation of algorithms in financial markets is the most advanced. Since 3 January 2018, Article 26 of the EU Markets in Financial Instruments Directive 2 (MiFID 2) requires investment firms to include details of the computer algorithms responsible for investment decisions and for executing transactions. On 12 February 2019, a motion for a European Parliament resolution was adopted in plenary session, following the adoption of the own-initiative report on 'A comprehensive European industrial policy on artificial intelligence and robotics'.

## Concluding remarks

The development of AI in a regulatory and ethical vacuum has triggered a series of debates on the need for its legal control and ethical oversight. AI-based algorithms that perform automated reasoning tasks

appear to control increasing aspects of our lives by implementing institutional decision-making based on big data analytics and have in effect rendered this technology an influential standard-setter.

The impact of existing AI technologies on the enjoyment of human rights, from freedom of expression, freedom of assembly and association, the right to privacy, the right to work, and the right to non-discrimination to equal protection of the law, needs to be carefully examined and qualified along with AI's potential to aggravate inequalities and increase the digital divide. In view of AI's potential to act in an autonomous manner, its sheer complexity and opacity, as well as the uncertainty surrounding its operation, make a comprehensive regulatory response essential to preventing the ever-expanding applications from causing social harm among a very heterogeneous range of individuals and social groups.

Such a response should entail the obligation for developers of AI algorithms to fully respect the human rights and civil liberties of all users by retaining uninterrupted human control over AI systems, address the effects of the emotional connection and attachment between humans and robots, and develop common standards against which a judicial authority that makes use of AI will be assessed. It should also focus on the allocation of responsibilities, rights and duties, and prevent the reduction of the legal governance process to a mere technical optimisation of machine learning and algorithmic decision-making procedures. Within this frame, new collective rights concerning data need to be introduced that will safeguard the ability to refuse to be subjected to profiling, the right of appeal and the right of explanation in AI-based decision-making frameworks.

Furthermore, legislators need to ensure that organisations which deploy and utilise these systems remain legally responsible for any damage caused and develop sustainable and proportionate informed consent protocols. Although [no law](#) can encode the entire complexity of technology as it is, let alone predict its future development, the EU could employ soft law instruments (such as regulatory technology assessments and real-time ethics audits) to anticipate or even shape technological trends, and ensure that disruptive technologies are deployed in a way that is consistent with the EU ethical *acquis*. Although some 'black swan' technological events will remain impossible to predict, an ethics-by-design approach could ensure that regulatory policy proactively adapts to an evolving disruptive ecosystem, and influences the design of technologies.

Given that AI systems are unstable objects of legal and ethical scrutiny, algorithmic impact assessments and audits may need to become a legal requirement. The 2017 European Parliament resolution on civil law rules on robotics – comprising a 'code of ethical conduct for robotics engineers', a 'code for research ethics committees', a 'licence for designers', and a 'licence for users' can serve as a governance model for a detailed process-based architecture of technology ethics in the AI field. The charter on robotics contained in the resolution combines an ex-ante ethics-by-design approach with a reflexive framing and a meta-ethical analysis of the governance process employed for the embedding of ethics into the structures for the development of this disruptive technology. This legislative initiative resolution should be part of a wider paradigm shift that could include the introduction of [new ethical principles](#) (such as the right not to be measured, related to possible misuses of AI and the internet of things, and the right to meaningful human contact, relating to possible misuses of care robots).

Given the difficulty and complexity of predicting the performance of many AI systems and their interactions, testing new forms of accountability and liability via novel methods of legal experimentation is more than necessary. EU legislators need to evaluate carefully whether there is a need for specific regulation related to AI-enabled decision-making and more importantly whether a [right of appeal and a right to redress](#) when AI is used for decisions affecting individuals should be introduced. Ethical auditing and prior algorithmic impact assessments need to become essential aspects of all efforts to control AI systems with built-in autonomy and self-learning capacities.

The work of the European Commission's High Level Expert Group on AI is worth being mentioned at this point: Their first draft of [Ethics Guidelines for Trustworthy AI](#) emphasises the need to ensure that AI is human-centric, developed, deployed and used with an 'ethical purpose', and recommends the

incorporation of specific requirements for trustworthy AI from the earliest design phase (accountability, data governance, design for all, governance of AI autonomy (human oversight), non-discrimination, respect for human autonomy, respect for privacy, robustness, safety, transparency), and highlights the importance of the auditability of AI systems, particularly in critical contexts or situations, as well as of ensuring a specific process for accountability governance.

The recently adopted [EP resolution on a Comprehensive European industrial policy on artificial intelligence and robotics](#) reinforced the European Parliament's focus on the need for the establishment of a guiding ethical framework and the political belief that Europe should take the lead on the global stage by deploying only ethically embedded AI. It recommends that EU Member States establish AI ethics monitoring and oversight bodies; encourage companies developing AI to set up ethics boards and draw up ethical guidelines for their AI developers; and requests an ethics-by-design approach that will facilitate the embedding of values such as transparency and explainability in the development of AI. All these considerations could constitute elements of a new EU-wide social contract on responsible innovation that might possibly place ethics-by-design at the epicentre of the technology development cycle. Such a contract could render anticipatory technology ethics tools fully operational and bring forward the role and limitations of ethical expertise as a source of epistemic authority that claims to represent the entirety of societal concerns.

## REFERENCES

[Artificial intelligence for Europe](#), Legislative Observatory (OEIL), European Parliament.

[Comprehensive European industrial policy on artificial intelligence and robotics](#), Legislative Observatory (OEIL), European Parliament.

Resolution of 27 January 2017, [civil law rules on robotics](#), on European Parliament.

[European coordinated plan on artificial intelligence](#), press release, Council of the EU, 18 February 2019.

[Draft AI ethics guidelines](#), European Commission, December 2018.

Communication on [Artificial intelligence for Europe](#), European Commission, 25 April 2018.

Communication on a [Coordinated plan on artificial intelligence](#), European Commission, 7 December 2018.

## DISCLAIMER AND COPYRIGHT

This document is prepared for, and addressed to, the Members and staff of the European Parliament as background material to assist them in their parliamentary work. The content of the document is the sole responsibility of its author(s) and any opinions expressed herein should not be taken to represent an official position of the Parliament.

Reproduction and translation for non-commercial purposes are authorised, provided the source is acknowledged and the European Parliament is given prior notice and sent a copy.

© European Union, 2019.

[stoa@ep.europa.eu](mailto:stoa@ep.europa.eu) (contact)

<http://www.europarl.europa.eu/stoa/> (STOA website)

[www.europarl.europa.eu/thinktank](http://www.europarl.europa.eu/thinktank) (internet)

<http://epthinktank.eu> (blog)

