**Author:**

**Dr. sc. Nediljka Rogošić**

**Artificial intelligence Ethics: questions and initiatives**

*Summary*

*For some time now, the business, professional and scientific world has been discussing the ethics of artificial intelligence and the ethical issues that arise or may arise during the application of artificial intelligence in various areas of human life.*

*Since the Study entitled Ethics and Artificial Intelligence: Issues and Initiatives was published in this regard, its content is presented in this paper, in order to draw attention and encourage reflection on the need and importance of defining the ethics of artificial intelligence. Recommendations were also given for the development of standards and drafting regulations at the national level, which is a prerequisite for reducing the negative impact of artificial intelligence on various areas of human life.*

# ***1. Introduction***

Organised by the Faculty of Philosophy and Religious Studies, University of Zagreb, Centre for Business Ethics, a Symposium was held in 2021 entitled “Artificial Intelligence, Economy and Business Ethics”. It was an opportunity to exchange knowledge and to stimulate the discussion about artificial intelligence and ethics as a very current and important topic.

This paper contains an overview of the results of a study called *Ethics and Artificial Intelligence: questions and initiatives (the Ethics of Artificial Intelligence: issues and Initiatives),* created by a group of authors from the University of West England at the request of the Panel for the Future of Science and Technology (STOA), managed by the Scientific Foresight Unit, within the European Parliamentary Research Service (EPRS) of the Secretariat of the European Parliament.[[1]](#footnote-1)

**The purpose of this paper is to encourage thinking about ethics and artificial intelligence, starting from the results presented in the mentioned study.**

The study deals with the ethical implications and moral issues arising from the development and application of AI. It contains six main chapters and a Summary and Addendum *Building Ethical Robots.*

The first and the introductory chapters address the question of AI (*artificial intelligence*, hereinafter referred to as AI) and what intelligence is, the definition of morality and ethics and their relationship with AI. It was also pointed out that artificial intelligence and machine learning bring enormous potential benefits, but that it is necessary to explore the full ethical, social and legal aspects of AI systems in order to avoid unintended, negative consequences and risks arising from the implementation of AI in society.

The second chapter entitled “Mapping the main ethical concerns and moral issues related to the introduction of AI” highlights the main ethical and moral issues related to the development and application of AI, which relate to their impact on society, human psychology, the financial system, the legal system, the environmental and planet impacts and the impact on trust.

The third chapter entitled “Ethical initiatives in the field of artificial Intelligence” presents international ethical initiatives, ethics violations and concerns faced by these initiatives. A number of case studies have also been described under this chapter.

The fourth chapter entitled “AI standards and regulations” deals with ethical standards and regulations that have been developed so far in the world or in certain national countries, while the fifth chapter provides an overview of national and international AI strategies. Chapter Six refers to emerging themes that present solutions to ethical issues through national and international strategies and solutions to governance challenges posed by artificial intelligence.

# ***2. Artificial intelligence, ethics and morality***

According to this study, *artificial Intelligence*[[2]](#footnote-2) (AI) refers to systems that show intelligent behaviour by analysing their environment and taking activities (actions) — with a certain degree of autonomy — to achieve specific objectives.

AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications)].[[3]](#footnote-3)

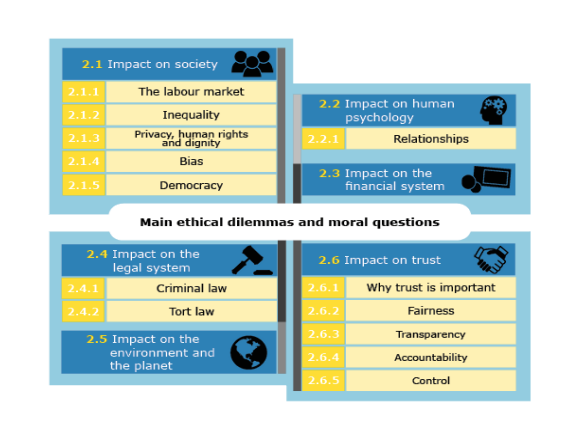
In addition to the definition of AI, the study also contains a definition of ethics and morals and clarifications regarding their relationship with AI. *Ethics* is defined as a set of moral principles that govern the behaviour of a person or the performance of an activity. *AI ethics is concerned with the important question of how human developers, manufacturers and operators should behave in order to minimise the ethical harms that can arise from AI in society, either arising from poor (unethical) design, inappropriate application or misuse.* The scope of AI ethics spans immediate, here-and-now concerns about, for instance, data privacy and bias in current AI systems; near- and medium-term concerns about, for instance, the impact of AI and jobs and the workplace; and longer-term concerns about the possibility of AI systems reaching or exceeding human-equivalent capabilities (so-called *superintelligence*). [[4]](#footnote-4)

In view of these circumstances, many initiatives have been launched and many ethical principles for robotics and AI have been published. The study provides that at least 22 different sets of ethical principles have been published since January 2017 and new ethical standards have emerged. (in particular the British Standards Institute and the IEEE Standards Association). In addition, and a growing number of countries (and groups of countries) have announced AI strategies (with large-scale investments) and set up national advisory or policy bodies.[[5]](#footnote-5)

# ***3. Main ethical concerns and moral issues related to the development and application of AI***

The main ethical concerns and moral issues related to the development and application of AI *(Diagram 1)* concern:[[6]](#footnote-6)

* Social impacts
* Psychological impacts
* Financial system impacts
* Legal system impacts
* Environmental impacts
* Impacts on trust

Graph 1 **Main ethical and moral issues related to development and implementation of AI**

Source: *The ethics of artificial intelligence: Issues and Initiatives*

According to this graphic, it can be concluded that the influence of AI on society and on trust is more detailed in relation to other areas, namely human psychology, the financial system, the legal system and the environment and the planet.

## ***3.1. Impact on society***

The impact on society encompasses the labour market, inequality, privacy, human rights and dignity, bias and democracy. Given the large and significant impact of AI on the society, as recognised in the study, below in Graph 2, a more detailed elaboration of the impact of AI on the society is given.

Graph 2 **Impact of AI on society**

**IMPACT OF AI ON SOCIETY**

Impact on economic growth and productivity

**IMPACT ON THE LABOUR MARKET**

Impact on the workforce

Labor market discrimination and effects on different demographics

Concentration of power among elites

Inequality: exploitation of workers

Benefit sharing

**IMPACT ON INEQUALITY**

Political instability

Privacy and data rights

Human Rights

Surveillance

**IMPACT ON PRIVACY, HUMAN RIGHTS AND DIGNITY**

Freedom of speech

**IMPACT ON BIAS**

Examples of AI Bias

Implications

**IMPACT ON DEMOCRACY**

News bubbles and echo chambers

The end of democracies

### *3.1.1. Impact on the labour market*

Regarding the impact of AI on the labour market[[7]](#footnote-7), concerns related to:

* the replacement of workers with technology (cessation of the need for some jobs, damage to the labour market)
* possible occurrence of mass unemployment
* technologies that will affect jobs, making many jobs obsolete or unnecessary.

The study stated that economists agree that AI will have a positive impact on the economy and productivity, while there are different opinions and surveys regarding the impact on the workforce, i.e. a significant number of economists agree on the positive impact of AI on the workforce, while also a significant number of researchers are concerned about the potential inequality of wages or personal earnings, unemployment, etc. In addition, discrimination may occur on the labour market, as the influence of AI will not be equally felt by all members of society, some are in greater danger and others are in lower danger. For example, those members of society who have fewer technical skills and know-how will face greater difficulties than those who have more technical know-how. There is a question of acquiring the knowledge and skills of young people. Those who have a sense and desire for science, technology, engineering, mathematics, etc. will not have difficulty working in a new environment while young people who do not have the same or similar desires may face significant difficulties in finding suitable jobs.

However*, while the development and application of AI and related technologies will require the creation of new jobs that will require specific knowledge and skills, there are legitimate concerns about potential negative consequences and the advantages and disadvantages of AI, i.e. threats to society and individual country populations, need to be identified in a timely manner at country level and strategic measures taken to take advantage of AI systems in the long term and to address or minimise negative impacts on society or the labour market.*

### *3.1.2. Impact on inequality*

The Study recognizes the concern about the possible emergence of inequality[[8]](#footnote-8) of workers on the labour market worldwide. Inequality is an ethical issue, but it is not included in discussions of AI ethics. However, the issue is important because inequality involves exploitation of workers, unequal sharing of benefits, concentration of power among “elites” and political instability. An example is given of how inequalities can arise in workplaces that require highly skilled workers, given that the jobs in these workplaces can be repetitive and boring, that there is a risk that the compensation for the work of these highly skilled workers will not be commensurate with the value of the final product, that individuals around the world who work in these jobs related to AI are most often not protected by employment contracts, that they do not know the work process (nor can they know it) as well as the value of the final product in the creation of which they participate, and more.

The study outlines *the potential benefits* for society and the ability of AI to tackle a number of the most serious global issues such as poverty, diseases, conflicts, and the like, which in the long run can have an impact on improving the quality of life of people around the world. In doing so, it is very important that these benefits be evenly distributed and accessible to more people. Also, companies and their owners can benefit from AI due to increased profitability, without this benefit being transferred to workers or transferred to a smaller number of workers, which potentially increases inequality in work and wages.

One of the problems highlighted in the study related to AI development is *the concentration of power* *in the hands of fewer individuals*. (which is already recognizable today as a consequence of the accumulation of technological, economic and political power concentrated in the hands of several “best players”).

If this power is not put in the service of the greater good or good for as many people around the world as possible, especially the poorest, it will lead to even greater inequality.

The problem may also arise due to the long—term large impact on public awareness formation due to “controlled” or selected information or foreclosure of competition from the market, as well as the problem regarding the collection, analysis and use of personal data.

The development of AI can also lead to *wealth inequality and political turmoil* and political polarisation. In doing so, personal beliefs and attitudes can be used to create an elite or special group that has power and gives individuals a special status in society. In terms of status and great influence, such a group is most often separated from other members of society, who, regardless of their knowledge and skills, lose the opportunity to participate and influence development and become unequal members of society, or are disadvantaged.

Every man should have the opportunity and the right to be realized as a person in the fullest sense of the word. Each individual or group of individuals should use their power to create a greater good for everyone, not just for the individual. And that is why the proposal to “consider AI a general good” makes sense.

*In order to reduce potential negative impacts from AI, it is necessary to adopt clear measures, policies and rules that should provide solutions for possible negative economic impacts, negative changes in work or available jobs, disparities in the allocation of benefits, technological surpluses of employees, i.e. unemployment and possible social and political tensions.*

### *3.1.3. Impact on privacy, human rights and dignity*

As mentioned above, the development of AI has raised several issues and sparked discussions on different areas of human life. This raises questions and raises debates about possible negative effects on the privacy, human rights and dignity of persons using AI. In order to prevent or minimise such impacts, “*The privacy and dignity of AI users must be carefully considered when designing service, care and companion robots, as working in people's homes means they will be privy to intensely private moments such as bathing and dressing, etc.”.*'[[9]](#footnote-9)

In addition, machine learning-based AI applications require access to large amounts of data. However, data subjects have limited rights in the way their data is processed and used (usually used to train models).

In addition to these concerns, ethical questions are raised, in particular, when it comes to the possibility that an individual may exercise control over data collected about him or her without his or her permission and which are used for purposes which are not known or acceptable to him or her and which may ultimately cause him or her harm of various proportions.

In this connection, it is necessary to answer the question of how to prevent the system from using data for purposes other than those for which the data is provided by the individual, or how to limit the system from compromising the privacy of people and the human rights of individuals directly or indirectly involved in the collection of personal data (for example family members)?

*These areas need to be regulated in order to prevent negative consequences, i.e. to ensure the protection of human rights, privacy and the dignity of individuals.*

### *3.1.4. Impact on bias*

One possible AI problem relates to bias that may arise as a result of data used to train the system or as a result of values held by system holders and users.

The problem of bias can have significant negative effects, as bias affects the lives of individuals, given that it gives some an unjustified advantage over others, it unjustifiably denies some and opens up opportunities for others and the like. This creates inequalities, inequalities or discrimination.

Bias implies inappropriate behaviour and treatment of others. And it should always comply with ethical principles that, when it comes to AI, should be clearly set. *This means that, in accordance with the abovementioned definition of* *AI ethics, principles should be laid down in order for developers, producers and operators to behave with a view to minimising the harm (in ethical terms) that may result from the application of AI in society, whether due to poor (unethical) design, inappropriate application or abuse.[[10]](#footnote-10)*

### *3.1.5. Impact on democracy*

The study pointed out that the concentration of technological, economic and political power in several mega corporations can enable their unnecessary influence on governments, but also the development and use of AI, which can threaten democracy in a certain way. For example, false or incomplete social media information can be used to shape, direct opinions, and manipulate voters, thus influencing the results of elections in a country. The study provides a number of such examples. However, for the development of democracy and trust in democracy, elections must be free and fair and without the possibility of manipulation.[[11]](#footnote-11) The study highlighted other threats to democracy with regard to the targeting of customized news to people according to their area of interest, which the system recognized (News bubbles and Echo chambers).

There is no doubt that the development and application of AI in the world are significant and beneficial, but care should be taken of the consequences that can arise for both individuals and society as a whole for various reasons.

According to *Kim Escherich*, Executive Innovation Architect: "*in this digitalised world, people frown on opinions that oppose the advancement and exploitation of technology. However, we must address the ethical aspects and consequences of its use. It is not certain that increasing automation, monitoring and the number of algorithms will help us make the right decisions. It may trigger the opposite; a greater sense of insecurity and far less human contact that can cause a decline in the quality of life of an individual. So we have to determine what comes first*: “smart” *societies or “*ethical*” societies? I'm for ethical societies. "*[[12]](#footnote-12)

## ***3.2. The impact of AI on human psychology***

Since man is increasingly communicating with the machine, the question arises as to what impact this will have *on real human relationships*. In addition to being of great help to robots, the study highlighted several dangers or threats such as the possibility of deception, manipulation, misuse of robots by their manufacturers, owners, or various hackers. Also, it is also possible that people could become psychologically dependent on robots, that robots are considered real person, that they give more attention and attention to robots than family members, friends and the like, loss of contact with people or withdrawal from everyday normal social relationships. In view of the above, there are concerns that AI may reduce our ability to work together or team together, weaken our emotional and broader human ties.[[13]](#footnote-13)

However, the question arises, can a machine replace a man in relationships with one another, or what will happen to a man who separates himself from normal relationships in the long run?

This question can be answered by Paul Tillich, who is considering the importance of human relations with one another. He states: “in *personal encounters, a person creates a personal relationship with another man in which each one becomes aware of their personal “I” and the other “you”. There is no natural object in the universe that could lead to that knowledge.”[[14]](#footnote-14) He who cannot establish the relation of "I" to "you" cannot establish a relation neither with truth nor good nor with the foundation of battle in which they are embedded. Justice in personal encounters means the fulfilment of every man's need to be realized both as “me” and as “you.”[[15]](#footnote-15)*

On the basis of this reasoning, it can be concluded that there can be no good separating man from everyday human relations, since this would mean changing his battle.

Tillich also points out that:[[16]](#footnote-16)

- A man can transcend himself in all directions of knowledge and mastery and no one can say where the limits of his power are.

- It can use anything to achieve its objectives, and it is limited only by its own limitation, which can be reduced to a minimum.

- By encountering the world, man is able to overcome any possible constraint, but there is one specific constraint that he constantly encounters, and that is another man. The other one, the “you”, represents a kind of wall that cannot be removed, used, or penetrated. Anyone who tries to do that destroys himself. *That "you" by its very existence demands to be recognized (and accepted) as "you" by our "I" and as "I" for itself. It is a need that is implicitly contained in the human struggle.*

We can conclude that it is in relationships between individuals that the dignity of every human is evident that cannot be achieved between man and machine. A machine cannot replace another human in that sense. Apart from the influence of AI on human relations, the study also highlighted the question of “personality”. The question was whether the need or the possibility of giving the machine/robot a “*personality”* should be considered, since the machine is increasingly taking on tasks and making decisions traditionally carried out by man. The legal status of robots and AI systems has been debated for a long time, but most ethical research on AI agrees that AI machines should not be given moral status, nor should they be considered a person.[[17]](#footnote-17)

## ***3.3. Impact of AI on the financial system***

The study looked at the impact of AI on the financial system through accountability and the possibility of market manipulation. The financial system is known to be one of the first areas where AI had a major impact, and estimates show that more than half of equity trading volumes are attributed to algorithms. The impact of AI on the financial system is summarised in Graph 3.

Graph 3 **Impact of AI on the financial system**

**IMPACT OF AI ON THE FINANCIAL SYSTEM**

Algorithmic trading can generate profits at a speed and frequency that is impossible for a human trader

Although today's autonomous agents operate within a relatively narrow scope of competence and autonomy, they nevertheless take actions with consequences for people.

**MARKET MANIPULATION**

Autonomic trading agents could also be used maliciously to destabilize markets, could also be used maliciously to destabilize markets or otherwise harm innocent parties

The increasing use of sophisticated machine learning techniques makes it difficult to understand

how they will perform in unanticipated circumstances.

The responsibility for trading algorithms rests with the organizations’ that develop and deploy them, autonomous agents may perform actions — particularly in unusual circumstances — that would have been difficult to anticipate by their programmers. Does that difficulty mitigate responsibility to any degree?

In this regard, theoretically, the agent could attempt to instigate arbitrage opportunities by taking malicious actions to subvert markets, for example by propagating misinformation, obtaining improper access to information, or conducting direct violations of market rules

**ACCOUNTABILITY**

Could an autonomous algorithm meet the legal definition of market manipulation, which requires 'intent'?

Wellmen and Rajan (2017) argue that trading agents will become increasingly capable of operating at wider levels without human oversight, and that regulation is now needed to prevent societal harm. However, attempts to regulate or legislate may be hampered by several issues

## ***3.4. Impact of AI on the legal system***

The next area of attention in the study is the legal system that AI machines can also have a major impact on, especially when it comes to criminal and civil law. The entire history of human laws was built on the assumption that humans, not robots, make decisions. It was pointed out that in a society where increasingly complex and important decisions are handed over to algorithms, there is a risk that existing legal frameworks will be insufficient when it comes to accountability. A summary of the impact of AI on the legal system is given in Graph 4.

Graph 4 **Impact of AI on the legal system**

**IMPACT OF AI ON THE LEGAL SYSTEM**

**LIABILITY** - Existing liability models may be inadequate to address the future role of AI in criminal activities

**PSYCHOLOGY -** There is a risk that AI robots could manipulate a user's mental state in order to commit a crime

**CRIMINAL LAW**

**TRADE, FINANCIAL MARKETS AND INSOLVENCY** - there are concerns that autonomous agents in the financial sector could be involved in market manipulation, price fixing and collusion. The lack of intention by human agents, and the likelihood that autonomous agents (AAs) may act together also raises serious problems with respect to liability and monitoring. It would be difficult to prove that the human agent intended the AA to manipulate markets, and it would also be difficult to monitor such manipulations.

**HARMFUL OR DANGEROUS MEDICINES** - In the future AI could be used by organised criminal gangs to support the trafficking and sale of

banned substances

**OFFENSES AGAINST A PERSON -** AI can generate more sophisticated fake content, new forms of harassment are possible

**SEXUAL OFFENSES -** There is a danger that AI embodied robots could be used to promote sexual objectification, sexual abuse and violence

**THEFT AND FRAUD, FORGERY AND IMITATION -** AI could be used to gather personal data, and forge people's identities. AI could also be used to commit banking fraud by forging a victim's identity, including mimicking a person's voice.

**TORT LAW** covers situations where one person's behaviour causes injury, suffering, unfair loss, or harm to another person. This is a broad category of law that can include many different types of personal injury claims.

**Tort laws** serve two basic, general purposes: 1) to compensate the victim for any losses caused by the defendant's violations; and 2) to deter the defendant from repeating the violation in the future.

**TORT LAW**

The question of the impact on the legal system is very complex and sensitive and needs to be addressed with great care.[[18]](#footnote-18)

*The impact of an AI on the legal system may be significant and provision should be made for possible liability and compensation for damage associated with the use of an AI, in particular in cases where the damage is not directly attributable to a person but to an AI (if the liability of an AI can be considered).*

## ***3.5. Environmental and planet impacts of AI***

The study highlights the impact of AI and robotic technology on the environment and the planet, especially in a situation where the world is facing major climate changes, given that AI requires a lot of energy, and robotic technology is associated with significant environmental pollution and the creation of a large amount of waste. Also, the increase in utilization of rare earth metals is significant. A summary of the impact of AI and robotics technology on the environment and the planet is given in Graph 5.[[19]](#footnote-19)

Graph 5

**Environmental and planet impacts of AI**

**ENVIRONMENTAL AND PLANET IMPACTS OF AI**

Increase the yield, and depletion rate of rare earth metals, degrading the environment further

**USE OF NATURAL RESOURCES**

At the end of their product cycle, electronic goods are usually discarded, leading to a build-up of heavy metals and toxic materials in the environment

**POLLUTION AND WASTE**

Increasing the production and consumption of technological devices such as robots will exacerbate this waste problem, particularly as the devices will likely be designed with 'inbuilt obsolescence' resulting in the generation of large amounts of electronic waste

Adoption of AI technology, particularly machine learning, will require more and more data to be processed. And that requires huge amounts of energy. AI will also require large amounts of energy for manufacturing and training

**ENERGY CONCERNS**

Alternatively, AI could actually help us take better care of the planet, by helping us manage waste

and pollution.

**WAYS AI COULD HELP THE PLANET**

*In order to address these problems and the risk of increased environmental pollution, rules and guidance should be adopted at national and global level to address these problems in a timely manner and to minimise risks.*

## ***3.6. Influence of AI on trust***

The Study highlighted that AI will change our daily lives in sectors such as transportation, service industries, healthcare, education, public safety and security, and entertainment, but that these systems must be introduced in ways that build trust and understanding and respect human rights. and civil rights (Dignum, 2018). They must follow fundamental human principles and values and protect the well-being of people and the planet. A summary of the impact of AI on trust is given in Graph 6. [[20]](#footnote-20)

Graph 6 **Influence of AI on trust**

**INFLUENCE OF AI ON TRUST**

To reach its full potential, we need to allow machines to sometimes just work all the time and make decisions on their own without human input; it is crucial to identify an effective way of trusting digital technologies so that we can harness their value, while simultaneously protecting fundamental rights and fostering the development of open, tolerant and just information societies.

**WHY TRUST IS IMPORTANT?**

It is essential that the decisions made by AI are fair and do not deepen already entrenched social inequalities, it is necessary to ensure that these decisions are free from bias and discrimination

**FAIRNESS**

An important aspect of fairness is knowing why the automated program made a certain decision

It should always be possible to find out why an autonomous system made a particular decision, especially if that decision caused harm

Other issues and problems with transparency include the fact that software and data are proprietary works, which means it may not be in a company's best interest to divulge how they address a particular problem. Many companies view their software and algorithms as valuable trade secrets that are absolutely key to maintaining their position in a competitive market.

**TRANSPARENCY**

Transparency also conflicts with privacy, as people involved in training machine learning models may not want their data, or inferences about their data to be revealed. In addition, the lay public, or even regulators may not have the technological know-how to understand and assess algorithms.

While transparency is often taken to mean the disclosure of source code or data, we don't have to see the computer source code for a system to be transparent, as this would tell us little about its behaviour. Instead transparency must be about the external behaviour of algorithms.

Accountability is one way to ensure trust in AI. Accountability ensures that if an AI makes a mistake or harms someone, there is someone that can be held responsible, whether that be the designer, the developer or the corporation selling the AI. In the event of damages incurred, there must be a mechanism for redress so that victims can be sufficiently compensated. One way of ensuring accountability is regulation. Winfield and Jirotka (2018) point out that technology is, in general, trusted if it brings benefits and is safe and well regulated. Their paper argues that one key element in building trust in AI is ethical governance – a set of processes, procedures, cultures and values designed to ensure the highest standards of behaviour. These standards of behaviour need to be adopted by individual designers and the organisations in which they work, so that ethical issues are dealt with as or before they arise in a principled manner, rather than waiting until a problem surfaces and dealing with it in an ad-hoc way.

**ACCOUNTABILITY**

Control is also one of the issues affecting public trust in AI. AI machines that make decisions with wider societal implications, such as algorithms that control millions of self-driving cars or news filtering algorithms that influence the political beliefs and preferences of millions of citizens, should be subject to oversight by society as a hole, requiring a 'society-in-the-loop' paradigm". As a way to address some of the threats of artificial intelligence, researchers have proposed ways to stop an AI system before it has a chance to escape outside control and cause harm.

**CONTROL**

How can we find the right level of trust? Taddeo points out that short-term shaping could play a key role in solving this problem. For example, pop-up messages alerting users to the results of an algorithmic search engine that have taken into account the user's internet profile or messages indicating that the impact/result of the algorithm may not be objective. However, infrastructure that applies standards such as fairness, transparency and accountability across sectors is needed in the long term.[[21]](#footnote-21)

*In order to create prerequisites for trust in AI, it is essential that AI makes fair decisions, that transparency is ensured, that levels of accountability are known and that there is an appropriate control system. However, more questions arise: can we trust artificial intelligence and what is the level of trust a person can show towards artificial intelligence? Could it be about trust in AI or is it an indirect speech about trust in those who developed or shaped AI? These issues should definitely be addressed by experts from different fields in order to avoid unwanted consequences in the future.*

# ***4. Ethical initiatives in the field of AI***

The study pointed out that many ethical considerations monitoring the development, use and effects of AI are known. These range from the potential effects AI could have on the fundamental human rights of citizens within a society to the security and utilisation of gathered data; from the bias and discrimination unintentionally embedded into an AI by a homogenous group of developers, to a lack of public awareness and understanding about the consequences of their choices and usage of any given AI. It was also pointed out that AI is based on previous revolutions in ICT and computing and, as such, will face a number of similar ethical problems.

Furthermore, it was pointed that although technology may be used for good, potentially it may be misused or it can be excessively anthropomorphised and humanised, blurring the lines between human and machine.

Given the major challenges, unresolved potential problems and dilemmas, unanswered questions about the meaning and impacts of AI systems and also the consequences of AI systems, many countries around the world (especially developed) have launched initiatives to regulate this area, offer acceptable solutions and answers specifically related to ethics and AI. International ethical initiatives are summarised below, and more can be found in the study as well as in many other published papers on the subject.[[22]](#footnote-22)

The study highlights that the formal legal framework is insufficient so far, but that several initiatives have been launched at a non-international level to investigate ethical concerns. The initiatives were launched by various institutes, associations, societies, partnerships or centres in the United States, Europe, Germany, the United Kingdom, Japan, the Netherlands, Belgium, Finland, Canada and Switzerland.[[23]](#footnote-23)

As regards the results of the work of these initiatives, the study concluded that all these initiatives agree that AI should be explored, developed, designed, deployed, monitored and used ethically – but each has different priority areas. The study also analysed and grouped these initiatives, according to the type of issues they want to address and some of the proposed approaches and solutions.

As indicated in the study, the initiatives give rise to a number of key issues that can be divided into several categories, as set out in Table 1.

Table 1

Categories of key issues addressed

international ethical initiatives

| *Number* | *Categories of key issues* | *Questions* |
| --- | --- | --- |
| 1. | Human rights and well-being | Is AI in the best interests of humanity and human well-being? |
|  |  |  |
| 2. | Emotional harm | Will AI degrade the integrity of the human emotional experience, or facilitate emotional or  mental harm? |
| 3. | Accountability and responsibility | Who is responsible for AI, and who will be held accountable for its actions? |
| 4. | Security, privacy, accessibility, and transparency | How do we balance accessibility and transparency with privacy and security, especially when it comes to data and personalisation? |
| 5. | Safety and trust | What if AI is deemed untrustworthy by the public, or acts in ways that threaten the safety of  either itself or others? |
| 6. | Social harm and social justice | How do we ensure that AI is inclusive, free of bias and discrimination, and aligned with public morals and ethics? |
| 7. | Financial harm | How will we control for AI that negatively affects economic opportunity and employment, and either takes jobs from human workers or decreases the opportunity and quality of these jobs? |
| 8. | Lawfulness and justice | How do we go about ensuring that AI - and the data it collects - is used, processed, and managed in a way that is just, equitable, and lawful, and subject to appropriate governance and regulation? What would such regulation look like? Should AI be granted 'personhood'? |
| 9. | Control and the ethical use – or misuse – of AI | How might AI be used unethically - and how can we protect against this? How do we ensure  that AI remains under complete human control, even as it develops and 'learns'? |
| 10. | Environmental harm and sustainability | How do we protect against the potential environmental harm associated with the development and use of AI? How do we produce it in a sustainable way? |
| 11. | Informed use | What must we do to ensure that the public is aware, educated, and informed about their use of and interaction with AI? |
| 12. | Existential risk | How do we avoid an AI arms race, pre-emptively mitigate and regulate potential harm, and ensure that advanced machine learning is both progressive and manageable? |

The study noted that overall, these initiatives aim to identify and design ethical frameworks and systems that establish benefits for people at the highest levels, prioritise both human society and the environment (without conflicting the two objectives) and mitigate AI-related risks and negative impacts, with a focus on ensuring AI accountability and transparency (IEEE, 2019).

In addition, the study stated that “*The ethically harmonised design of IEEE[[24]](#footnote-24): a vision for prioritising the well-being of people with autonomous and intelligent systems*” (V 1; 2019) is one of the most important documents published so far, on ethical issues that AI can raise and various proposed ways to mitigate them.

In addition, the *AI section of the study describes* *case studies,* such as: robots in health care, autonomous vehicles, warfare and weapons.[[25]](#footnote-25)

# ***5. AI standards and regulations***

As indicated in the study, standards that exist are still under development and limited information is publicly available.

The study lists BS 8611 – *Guide for the* *Ethical Design and application of robots and robotic systems* (British Standard BS 8611, 2016) as one of the possibly earliest developed AI ethical standards. It is based on 20 different ethical threats and risks, grouped into four categories: social, application, commercial, financial and environmental.[[26]](#footnote-26)

In addition, *the IEEE standards Association* has also launched a global initiative to develop *the ethics of Autonomous and intelligent systems*. There are currently 14 working groups on IEEE standards working to develop AI standards.[[27]](#footnote-27)

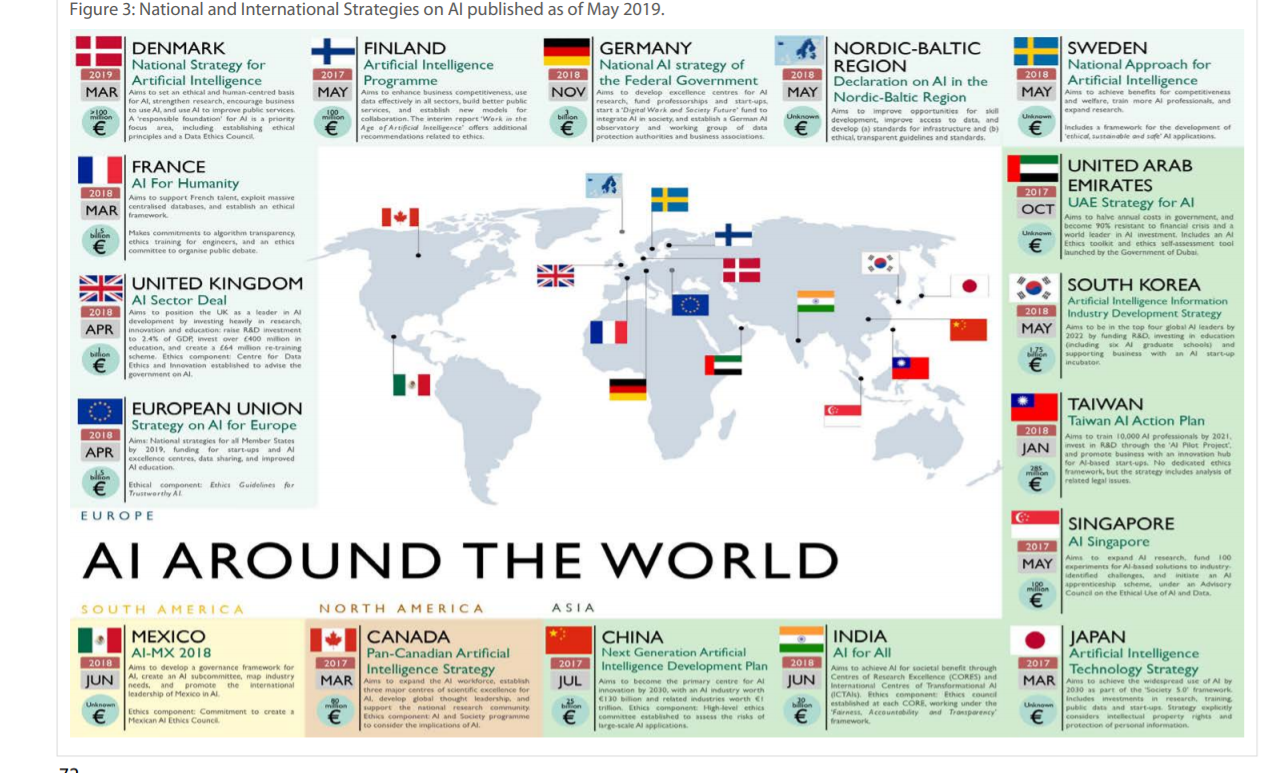
# ***6. International and national AI strategies***

According to the study, the first National AI Strategy was launched by Canada in March 2017 and soon thereafter by the technological leaders of Japan and China. In Europe, the European Commission adopted *a Communication on artificial Intelligence*, initiating the development of independent strategies of each Member State, while the USA Intelligence Initiative is expected soon.

Also, the study noted that Russia is making intensive efforts to formalize its ten-point AI plan. Figure 1 shows the countries that have developed national strategies on AI.

Figure 1

**National and international AI strategies**



Source: *The ethics of artificial intelligence: Issues and Initiative*

***What has been done in the Republic of Croatia?***

The Centre for artificial Intelligence (CAI) operates in the Republic of Croatia. It is the leading research centre for AI and was established in October 2019.[[28]](#footnote-28) The CroAI Association is also active in promoting AI-based artificial intelligence and solutions, established in February 2020.

In addition, in February 2021, the National Development Strategy of the Republic of Croatia was adopted until 2030 and is within the Strategic goal 11. The “Digital transition of society and economy” stated “*particular attention will be paid to improving the legislative framework regulating the use and privacy of data in order to strengthen opportunities for digital entrepreneurship, including the field of Big data analysis and artificial intelligence. When improving the legislative framework, the impact of artificial intelligence and analysis of “Big data” on fundamental human rights will also be considered and protection against discrimination ensured”*.[[29]](#footnote-29)

# ***Conclusion***

The rapid development of technology and artificial intelligence has significantly changed human life and brought significant benefits. However, in practice, certain problems and risks and potential negative impacts on society, human psychology, the financial system, the legal system, the environment and the planet as well as trust in AI have also been identified, in particular as there are no or only pending regulations and standards that would more clearly regulate the conditions and responsibilities associated with the development and application of AI. Many countries have already joined the development of National AI strategies and regulations, while ethical standards and regulations on AI are under development and are not accessible to the general public. In the National Development Strategy, the Republic of Croatia stressed the need to adopt a legislative framework that would, among other things, regulate the issue of the development and application of artificial intelligence.

When adopting the legislative framework and ethical standards on AI, the results of this study can be of great help to the Republic of Croatia and other countries, since it contains a systematic overview of problems and risks and possible negative impacts on different areas of human life. The study can encourage EU member States to take this problem seriously, to form working groups that will bring together experts from different fields who will approach the development of standards and drafting regulations wisely and prudently.

# Literature

The ethics of artificial intelligence: Issues and Initiatives, study, Panel for the Future of Science and Technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>

Artificial Intelligence Centre, [https://cai.fer.hr/#](https://cai.fer.hr/)

National Development Strategy of the Republic of Croatia until 2030, approach,

<https://narodne-novine.nn.hr/clanci/sluzbeni/2021_02_13_230.html>, available 19.7.2021

[Kim ESCHERICH](https://www.ibm.com/blogs/nordic-msp/author/kim-escherich/)**,** Executive Innovation Architect, *The ethical quest for A smart Society,* <https://www.ibm.com/blogs/nordic-msp/ethical-quest-smart-society/>

Paul TILLICH; *Love, Power and Justice*

1. *The ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the Future of Science and Technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, Page I, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>, access to: 18. 1. 2021. [↑](#footnote-ref-1)
2. Simply put, intelligence is “the ability to acquire and apply knowledge and skills and to manage one's own environment,” *the ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the Future of Science and technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, Page 1, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>, access to: 18. 1. 2021. [↑](#footnote-ref-2)
3. *The ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the Future of Science and Technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, Page 1, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>, access to: 18. 1. 2021. [↑](#footnote-ref-3)
4. *The ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the Future of Science and Technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, p. 2-3, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>, access to: 18. 1. 2021. [↑](#footnote-ref-4)
5. *The ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the Future of Science and Technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, Page 3, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>, access to: 18. 1. 2021. [↑](#footnote-ref-5)
6. Compare: *The ethics of artificial intelligence: Issues and Initiatives,* Study, Panel for the Future of Science and Technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, Page 5, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>, access to: 18. 1. 2021. [↑](#footnote-ref-6)
7. *The ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the Future of Science and technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, Page 6-8, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>, access to: 18. 1. 2021. [↑](#footnote-ref-7)
8. *The ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the Future of Science and Technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, Page 8-12, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>, access to: 18. 1. 2021. [↑](#footnote-ref-8)
9. *The ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the future of Science and technology, European Parliamentary Research Service, Scientific focus Unit (STOA), March 2020, Page 12-15, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>, access to: 18. 1. 2021. [↑](#footnote-ref-9)
10. Compare: *The ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the Future of Science and Technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, Page 2, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>, access to: 18. 1. 2021. [↑](#footnote-ref-10)
11. *The ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the Future of Science and Technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, Page 16-18, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>, access to: 18. 1. 2021. [↑](#footnote-ref-11)
12. # [Kim ESCHERICH](https://www.ibm.com/blogs/nordic-msp/author/kim-escherich/)**,** Executive Innovation Architect, *The Ethical Quest for a Smart Society,* available at: <https://www.ibm.com/blogs/nordic-msp/ethical-quest-smart-society/>, access to: 16. 7. 2021.

    [↑](#footnote-ref-12)
13. Compare: *The ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the Future of Science and Technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, Page 18-20, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>, approach, 16. 7. 2021. [↑](#footnote-ref-13)
14. Paul TILLICH; *love, power and Justice*, Page 78. [↑](#footnote-ref-14)
15. Paul TILLICH; *love, power and Justice*, Page 78. [↑](#footnote-ref-15)
16. Paul TILLICH; *love, power and Justice*, Page 78. [↑](#footnote-ref-16)
17. Compare: *The ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the future of Science and technology, European Parliamentary Research Service, Scientific focus Unit (STOA), March 2020, Page 27-22, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf> [↑](#footnote-ref-17)
18. See more about this: *The ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the Future of Science and Technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, pages 22-28, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf> [↑](#footnote-ref-18)
19. Compare: *The ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the Future of Science and Technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, pages 28-29, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf> [↑](#footnote-ref-19)
20. Compare: *The ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the Future of Science and Technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, pages 29-36, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf> [↑](#footnote-ref-20)
21. *The ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the Future of Science and Technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, Page 30, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>, access to: 18. 1. 2021. [↑](#footnote-ref-21)
22. More about this: *The ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the Future of Science and Technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, Page 37-65, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>, access to: 18. 1. 2021. [↑](#footnote-ref-22)
23. *The study provided a table containing information on the initiative - the originator, the originator's country of origin, key issues addressed by the initiative, publications and funding sources., the ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the Future of Science and Technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, Page 38-41, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>, access to: 18. 1. 2021. [↑](#footnote-ref-23)
24. Eye-triple-E, signifies the Institute of Electrical Engineers and electronics. [↑](#footnote-ref-24)
25. For more information on international initiatives, ethical damage and concerns addressed by international initiatives, see: *The ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the Future of Science and Technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, Page 37-65, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>, access to: 18. 1. 2021. [↑](#footnote-ref-25)
26. More about the above standards in: *The ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the Future of Science and Technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, Page 66, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>, access to: 19. 1. 2021. [↑](#footnote-ref-26)
27. More about the above standards in: *The ethics of artificial intelligence: Issues and Initiatives,* study, Panel for the Future of Science and Technology, European Parliamentary Research Service, Scientific Foresight Unit (STOA), March 2020, Page 66-67, available at: <https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf>, access to: 19. 1. 2021. [↑](#footnote-ref-27)
28. AI Hub: available at: [https://cai.fer.hr/#](https://cai.fer.hr/), access: 19. 7. 2021. [↑](#footnote-ref-28)
29. N National Development Strategy of the Republic of Croatia until 2030, available at:

    <https://narodne-novine.nn.hr/clanci/sluzbeni/2021_02_13_230.html>, pristup:19. 7. 2021. [↑](#footnote-ref-29)