Miskolci Jogtudó HU ISSN 2630-9505
2022/3. szám, 45–58.

EXAMINATION OF ETHICAL ISSUES REGARDING INTELLIGENT CARS[*]

Ibolya Stefán[**]

Abstract

The appearance and application of artificial intelligence, IoT devices, intelligent cars as the results of digitalisation, and the Fourth Industrial Revolution pose an unprecedented challenge for humanity as these technologies fundamentally change our current knowledge. The European Union recognised the potential of the novelties very early on. Therefore, it has issued several documents, studies, recommendations, and guidelines in recent years. This paper aims to examine the ethical issues arising from the use of smart cars. In this regard, we present the various levels of automatisation of intelligent vehicles, then, the relevant parts of two studies, '*Ethics Guidelines for Trustworthy AI'* – published by the Independent, High-level Expert group –, and '*Ethics of Connected and Automated Vehicles'* – prepared by the Expert Group of European Commission – under the discussed topic. During the examination, we focus on civil liability issues; in this respect, we also cover the phenomena of 'Human-In-The-Loop' and 'Human-Out-of-The-Loop'.

Keywords: intelligent cars, artificial intelligence, ethics, HITL, HOTL

1. Introduction to 'intelligent' and self-driving cars

1.1. Conceptual basics

A thorough analysis of the subject requires a brief explanation of the technology and a description of the categories of self-driving and intelligent vehicles. Before this, it is necessary to clarify the importance of artificial intelligence – hereinafter referred to as 'AI' – regarding the mentioned vehicles. Describing AI is quite challenging because there are several concepts in the literature, which are focusing on different attributes of the novelty. The proposal for Artificial Intelligence Act defines the technology as follows: *"artificial intelligence system (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with"*.[1]

[*] "Supported by the ÚNKP-21-3 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund."

[**] Ibolya STEFÁN, PhD student, University of Miskolc, Deák Ferenc Doctoral School of Law, ibolya.stefan@uni-miskolc.hu. Supervisor: Dr. Réka Pusztahelyi PhD, Associate Professor, Department of Civil Law, Faculty of Law, University of Miskolc.

[1] Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, European Commission, Brussels, 21.4.2021. (COM(2021) 206 final).

Considering intelligent vehicles[2], the Society of Automotive Engineers' – hereinafter referred to as 'SAE' – classification should be mentioned. The SAE document distinguishes six levels regarding intelligent cars.

Level 0 – *Driving is not automatised;* this category includes vehicles with no driver assistance system, and the vehicle's human driver performs all driving functions.

Level 1 – *Driving Assistance* and Level 2 – *Partial Driving Automation* involves vehicles with driver assistance function – e.g. lane-keeping –, which assist but do not drive the car.

Level 3 – *Conditional Driving Automation* means vehicles capable of unsupervised driving in a limited number of driving situations but still have a driver who can interfere. The driver must take back control of the car if it is necessary.

Level 4 – *High Driving Automation* indicates that only a few driving modes are automated; control or supervision is not required from the operator. In a technical sense, the vehicle drives itself when the automatic driving option is enabled.

Level 5 – *Full Driving Automation* means the vehicle is fully automated, and an individual's control or supervision is not required.[3]

In light of this classification, all vehicles equipped with AI can be considered intelligent; however, only vehicles on Levels 4 and 5 can be categorised as self-driving. Therefore, in this study, we use the terms 'intelligent car' and 'intelligent vehicle' as they are broader than 'self-driving cars'.

Intelligent cars are getting more recognised in the past few years, as they have several benefits and can be utilised in several fields. Intelligent vehicles are used primarily for the transportation and transfer of goods.

Regarding transportation, there are intelligent taxis, driverless buses, and automated metro systems. In February 2020, California Public Utilities Commission issued Drivered Deployment permits to two companies to use autonomous vehicles on selected public roads in San Francisco between 22 and 6 hours, up to 30 mi/h speed, only if a safety driver is also present in the vehicle. The companies may collect fares for the service and offer shared rides.[4] Driverless buses operate in several cities worldwide. Another popular transportation form is driverless buses or autonomous shuttle buses. In 2021, France was the first country in the world, that approve the operation of driverless buses – SAE Level 4 – without a human driver on board.[5] Automated metro systems are also widespread; several GoA4 operate without a human driver on board globally, e.g. Paris. *"GoA4 reduces response times and variabilities that come from human control, which optimises the running time of trains by increasing the average speed and reducing dwell time in stations. This enhances reliability, shortens waiting times for*

https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF (Date of download: 25.06.2022.).

[2] Intelligent vehicles consist of software and hardware elements; AI appears in cars as ADS – an advanced driving system – essentially the technology that controls the car.

[3] Society of Automotive Engineers: Surface Vehicle Recommended Practice – Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (J3016_202104). https://www.sae.org/standards/content/j3016_202104/preview/ (Date of download: 25.06.2022.).

[4] California Public Utilities Commission: CPUC Issues First Autonomous Vehicle Drivered Deployment Permits. https://www.cpuc.ca.gov/news-and-updates/all-news/cpuc-issues-first-autonomous-vehicle-drivered-deployment-permits (Date of download: 15.11.2022.).

[5] Tom BATEMAN: France approves fully autonomous bus for driving on public roads in a European first. https://www.euronews.com/next/2021/12/01/france-approves-fully-autonomous-bus-for-driving-on-public-roads-in-a-european-first (Date of download: 15.11.2022.).

passengers and reduces operational costs. Train regulation systems can also optimise energy use and reduce maintenance costs through smoother driving standards and optimal acceleration and braking profiles."[6]

Considering the transfer of goods, automatic guided or robotic vehicles (AGVs) have been around for decades in different forms and levels of automatisation. Nowadays, these unmanned machines are able to transport goods in the facility with the help of floor markers, magnets, lasers and sensors. As the technology evolved quickly, the number of automated mobile robots (AMRs) will also rise soon. AMRs are similar to AGVs; the significant difference is that only the destination is programmed into AMRs, but not a specific path; AMRs determine this.[7] As we face driver shortage and a growing number of goods, the future of freight transport lies in self-driving trucks. It has yet to become a reality, as the operation of these trucks requires a driver on board. However, several brands are testing self-driving options on public roads to realise this goal.[8]

1.2. Methods of machine decision-making

The applied methods should also be examined in machine decision-making to learn how intelligent cars work. The techniques can be divided into two groups and six categories, as follows:

1. *Classical Methods*
   a) *Rule-Based Methods* rely on a rule database based on traffic laws, driving experience and knowledge; the applied strategies are established concerning various vehicle statuses. Unfortunately, this method is less robust and unable to manage complex requests.
   b) *Optimisation Methods* often use reward and utility functions to make a decision, e.g. Model Predictive Control or Game-Theory. The latter is a well-known theory among researchers studying AI and intelligent vehicles. According to this, the 'ideal strategy' is assumed to be adopted by all agents, and then behaviours are developed in accordance with the complementary strategies of other agents. The disadvantage of this method is that the ideal strategy and practice are inconsistent.
   c) *Probabilistic Methods* apply the mathematic probability theory to model action, and it can be used with other models; however, it does not give the best results in a complex environment.
2. *Learning-Based Methods*
   a) *Statistic Learning-Based Methods* are based on data training, which results in human-like decision-making, e. g. SVM and AdBoost. This method can be used in simple scenarios. It requires a large amount of data, while the outcomes are less appropriate.

[6] GoA4: The Way Forward for Metro Systems Worldwide. https://www.wsp.com/en-gb/insights/goa4-the-way-forward-for-metro-systems-worldwide (Date of download: 15.11.2022.).
[7] Autonomous Vehicles and the Increasing Demand for Automated Material Handling. https://www.kellertechnology.com/blog/autonomous-vehicles-and-the-increasing-demand-for-automated-material-handling/ (Date of download: 15.11.2022.).
[8] Daimler Trucks begins testing automated trucks on public roads. https://torc.ai/daimler-trucks-begins-testing-automated-trucks-on-public-roads/ (Date of download: 15.11.2022.).

b) *Deep Learning-Based Methods* are similar to machine learning, but the former is learning through neural networks. Likewise, the previous method also uses a massive amount of data[9], but the results are more accurate.
c) *Reinforcement Learning-Based Methods* are often used for decision-making. By experimenting with different behaviours, reinforcement learning aims to teach agents how to maximise returns. Agent actions may then be modified following reward functions. Models may be updated and iterated repeatedly using existing and new data.[10]

Consequently, none of the above-mentioned methods is suitable for intelligent cars, as they cannot model human-like decision-making properly.

Concerning decision-making, the Markov Decision Process should also be mentioned, which is *"… one of the common methods to describe the human-like decision-making process. Based on MPD, a kind of human longitudinal decision model is designed to solve the speed planning problem at a signalised intersection However, this model does not consider horizontal decision-making…the Hidden Markov Model (HMM) is used to describe the lane changing intention…the Partial Observable Markov Decision Process (POMDP) is used to make decisions at unsignalized intersections, realising human-like driving."*[11]

A new method, the inverse reinforcement model, should also be mentioned. *"IRL refers to an algorithm that reversely deduces the reward function, on the premise of providing a strategy or an expert example, so that the agent can learn how to make decisions on complex issues, through expert demonstration."*[12] As stated in the study of Wu et al., better results can be achieved with a new method which combines the inverse reinforcement model and the previously mentioned semi-Markov model. The new method is cardinal as it is more stable than the classic methods, learns from human drivers and is more accurate to human driving behaviours. Therefore, this method can be the key to safe, intelligent vehicles; however, there are some errors regarding its application.[13]

Participation of individuals in the learning method is a crucial question these days. Human-in-the-Loop refers to the learning process when people are involved in training data, algorithms and machine learning models. Participating individuals provide feedback to the algorithm on whether its decision (output) is correct or not. As a result, if more people use the technology, it will be more accurate. While this method may seem time-consuming, more training helps make it faster, tailored and transparent.

Moreover, this method helps to improve the novelty to be more ethical and fair, as it learns from humans. The application of human-in-the-loop makes it possible for the algorithm

[9] Intelligent cars are equipped with many sensors and cameras that help to collect information. The building of intelligent vehicles both technically – hardware – and technologically – software – is very complicated. See more on the topic, Hong CHENG: *Autonomous Intelligent Driving*, Springer-Verlag, London, 2011; Thor I. FOSSEN – Kristin Y. PETTERSEN – Henk NIJMEIJER: *Sensing and Control for Autonomous Vehicles*, Springer, 2017; Markus MAURER – Christian J. GERDES – Barbara LENZ – Hermann WINNER: *Autonomous Driving*, Springer, Berlin, 2016; Daniel WATZENIG – Martin HORN: *Automated Driving*, Springer, Cham, 2017.
[10] Qi LIU – Xueyuan LI – Shihua YUAN – Zirui LI: Decision-Making Technology for Autonomous Vehicles: Learning-Based Methods, Applications and Future Outlook, *2021 IEEE International Intelligent Transportation Systems Conference*, 2–5. https://www.semanticscholar.org/paper/Decision-Making-Technology-for-Autonomous-Vehicles%3A-Liu-Li/d4717c32fa6463403119341079bc9225a3055537#citing-papers (Date of download: 10.06.2022.).
[11] Zheng WU – Fangbing QU – Lin YANG – Jianwei GONG: Human-like Decision Making for Autonomous Vehicles at the Intersection Using Inverse Reinforcement Learning, *Sensors*, 2022(22), 2.
[12] WU – QU –YANG – GONG: op. cit., 2–3.
[13] WU – QU –YANG – GONG: op. cit., 17–18.

to learn good practices and later use them in similar situations. On the contrary, human-out-of-the-loop systems do not rely on the input of individuals, as the technology collects and trains data on its own. People evaluate it only in the last stage.[14]

Our perspective is on applied technologies that individuals should take a significant part in algorithm training, personalisation and evaluation. The system should learn good practices, patterns and apply these in accident scenarios – as described above – to make an ethical, safe and human-like decision, which is cardinal in road traffic. We believe it is also essential to ensure human drivers the possibility of human oversight to interfere in the operation of intelligent cars to avoid any failures and accidents on the road.

2. Regulatory framework of intelligent cars

Regulation of emerging technologies, including artificial intelligence and intelligent vehicles, is becoming more and more crucial. Several resolutions, agreements, and other documents have been published in recent years.

The *'Convention on Road Traffic'* (Vienna, 8 November 1968) – hereinafter referred to as the 'Vienna Convention' – is significant regarding the international regulation of road traffic. The Vienna Convention establishes standards for drivers, road signs, and many other aspects of road traffic. Following the main rules, a driver is essential for any moving vehicle or combination of vehicles. Additionally, physical and mental ability, as well as a fit physical and mental state, are required as the driver may be capable of driving the car at all times[15]. The revision of Articles 1 and 34bis of 14 January 2021 – which has come into force on 14 July 2022 – in response to intelligent technologies in road traffic, defines autonomous driving and establishes operational standards.[16]

The European Union also started to examine the issues of intelligent vehicles and created many initiatives and strategies to address the difficulties of regulating the novelties. A single European framework should be established, along with cooperation on connected and automated driving, as well as assurances of privacy and data protection, security, and public awareness, according to the *'Amsterdam Declaration'*, which was adopted by the European Transport Ministers on 14 April 2016.[17] As a result, *'Communication on a European strategy for cooperative intelligent transport systems'* was published by the European Commission on

[14] Robert KOCH: Human in the Loop: The Human in the Machine. https://www.clickworker.com/customer-blog/human-in-the-loop-ml/ (Date of download: 12.06.2022.).
[15] Vienna Convention on Road Traffic of 1968, Article 8.
[16] *"The requirement that every moving vehicle or combination of vehicles shall have a driver is deemed to be satisfied while the vehicle is using an automated driving system which complies with: (a) domestic technical regulations, and any applicable international legal instrument, concerning wheeled vehicles, equipment and parts which can be fitted and/or be used on wheeled vehicles, and (b) domestic legislation governing operation. The effect of this Article is limited to the territory of the Contracting Party where the relevant domestic technical regulations and legislation governing operation apply."* Vienna Convention on Road Traffic of 1968, Article 34bis.
[17] Declaration of Amsterdam on Cooperation in the Field of Connected and Automated Driving, 2016. https://www.regjeringen.no/contentassets/ba7ab6e2a0e14e39baa77f5b76f59d14/2016-04-08-declaration-of-amsterdam---final1400661.pdf (Date of download: 25.06.2022.).

30 November 2016, focusing on the most critical problems, including interoperability, data protection, and cyber-security.[18]

Considering the Hungarian legislation on intelligent vehicles, it is essential to mention the *Law-Decree Nr. 3 of 1980 on Promulgation of 1968 Convention on Road Traffic*, as it ratified the above-mentioned Vienna Convention. *Act 1 of 1988 on Road Traffic* outlines the fundamental principles of road transport, as well as the rights and responsibilities of the participant. *Joint Decree Nr. 1/1975 (II.5) of the KPM and BM on the Rules of Road Traffic* provides the main rules of road traffic, such as the definition of the driver, the rules for safe driving, vehicle traffic, or the regulations in the event of technical malfunctions.

As regards the regulation of highly automated driving system testing on test tracks and open public highways, Decree Nr. 5 of 1990 (IV. 12.) *on Technical Inspection of Road Transport Vehicles* and Decree Nr. 6 of 1990 (IV. 12.) *on Technical Requirements for the Placing into Circulation and Maintenance in Circulation of Road Transport Vehicles* should be mentioned, both issued by the Minister for Transport, Communication and Construction. Decree Nr. 5 of 1990 (IV. 12.) on Technical Inspection of Road Transport Vehicles differentiates non-autonomous and autonomous vehicles for development purposes.

3. Ethical documents regarding new technologies

3.1. The United Nations Educational, Scientific and Cultural Organization

The United Nations Educational, Scientific and Cultural Organization – hereinafter referred to as 'UNESCO' – issued *'Recommendation on the Ethics of Artificial Intelligence'*. Among many others, the document aims to provide values and principles to States regarding the preparation of their legislation and policies on AI. The recommendation lists the following values:

1. Respect, protection and promotion of human rights and fundamental freedoms and human dignity,
2. Environment and ecosystem flourishing,
3. Ensuring diversity and inclusiveness,
4. Living in peaceful, just and interconnected societies.

The document also catalogues essential principles, such as:

1. Proportionality and do no harm,
2. Safety and security,
3. Fairness and non-discrimination,
4. Sustainability,
5. Right to privacy and data protection,
6. Human oversight and determination,
7. Transparency and explainability,
8. Responsibility and accountability,
9. Awareness and literacy,

[18] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A European strategy on Cooperative Intelligent Transport Systems, a milestone towards cooperative, connected and automated mobility, European Commission, Brussels, 30.11.2016, (COM(2016) 766 final). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016DC0766&from=EN (Date of download: 25.06.2022.).

10. Multi-stakeholder and adaptive governance and collaboration.

As we look at the list of values and principles, the focal point of the recommendation is universal, equal human rights that must be respected and considered both in the phase of technology design and its application.

Furthermore, the document highlights several policy areas:

1. *Ethical impact assessment*: the document sets out for Member States the introduction of impact assessment to identify the benefits and risks of AI systems, their impact on human rights and assurance methods (risk prevention, mitigation measures). Governments should establish a regulatory framework that outlines a procedure – especially for public authorities – for conducting ethical impact analyses on AI systems in order to foresee outcomes, reduce risks, prevent negative outcomes, encourage citizen participation, and address societal challenges.
2. *Ethical governance and stewardship*: Member States should guarantee that AI governance instruments are transparent, harms caused by the application of AI systems are investigated, and proper remedial actions are provided; develop appropriate regulatory frameworks to ensure accountability, responsibility, transparency and explainability regarding AI systems.
3. *Data policy*: Member States should protect the right to privacy and establish proper legislation and data policies considering the technology. [19]
4. *Development and international cooperation*: *"Member States should promote AI ethics research by engaging international organizations and research institutions, as well as transnational corporations, that can be a basis for the ethical use of AI systems by public and private entities, including research into the applicability of specific ethical frameworks in specific cultures and contexts, and the possibilities to develop technologically feasible solutions in line with these frameworks."*[20]
5. *Environment and ecosystems*: in accordance with the recommendation Member States should ensure the data, energy, and resource-efficient AI systems.
6. *Gender*: AI system should help to achieve gender equality, it must be ensured that women and girls' human rights and fundamental freedoms are not violated by the technology.
7. *Culture*: *"Member States are encouraged to incorporate AI systems, where appropriate, in the preservation, enrichment, understanding, promotion, management and accessibility of tangible, documentary and intangible cultural heritage, including endangered languages as well as indigenous languages and knowledges, for example by introducing or updating educational programmes related to the application of AI systems in these areas, where appropriate, and by ensuring a participatory approach, targeted at institutions and the public."*[21]
8. *Education and research*: Members States should cooperate with international organisations and educational institutions to ensure 'AI literacy education'.
9. *Communication and information*: Member States should promote access to information and knowledge through AI (e.g. academia, journalists and researchers).

[19] UNESCO: Recommendations on the Ethics of Artificial Intelligence, 2021, 10–30. https://unesdoc.unesco.org/ark:/48223/pf0000381137 (Date of download: 13.09.2022.).
[20] UNESCO: op. cit., 30.
[21] UNESCO: op. cit., 32.

10. *Economy and labour*: in accordance with the recommendation, Member States should assess the impact of AI in the labour market and support research projects to analyse this relation.[22]
11. *Health and social well-being*: *"Member States should endeavour to employ effective AI systems for improving human health and protecting the right to life, including mitigating disease outbreaks, while building and maintaining international solidarity to tackle global health risks and uncertainties, and ensure that their deployment of AI systems in health care be consistent with international law and their human rights law obligations."*[23]

Consequently, the document has several values and principles that can be utilised regarding regulating AI and intelligent cars, focusing on ethical issues.

3.2. The Institute of Electrical and Electronics Engineers[24]

In this section, we briefly highlight the guidelines set out in the IEEE's 'Ethically Aligned Design – A Vision for Prioritising Human Well-being with Autonomous and Intelligent Systems' (hereafter referred to as 'EAD'). It should be noted that the terminology used in the policy document is not AI, but A/IS – which means Autonomous and Intelligent Systems – the concepts are the same in practice, which is important regarding our topic, as technology is an essential part of intelligent vehicles. The international document, which shares many similarities with the EU legislation – while the documents have different scopes –, also sets out Guidelines and general principles which are binding, and particular attention should be paid to these principles in the design process:

1. Technology shall be created and used with respect for protecting and promoting *human rights*.
2. Developers must accept increasing well-being as the primary criterion for development.
3. Based on the principle of *data security*, designers should, among other things, ensure that individuals have access to their data which should be shared securely.
4. Under *effectiveness*, developers must demonstrate the technology's suitability for the purpose and demonstrate its effectiveness.
5. *Transparency* is the basis for decision-making.
6. According to the principle of *accountability*, there must be a clear rationale and justification behind each decision made by the technology.
7. *Awareness of misuse* means that designers should protect against the abuse and risks of using technology.
8. The requirement for *competence* defines the active role of developers by making it their responsibility to define the knowledge and skills required for operation.[25]

While the document does not involve specific rules regarding intelligent vehicles, the guidelines mentioned earlier are fundamental findings on the ethical guidelines of autonomous

[22] UNESCO: op. cit., 33–37.
[23] UNESCO: op. cit., 37.
[24] Hereinafter referred to as 'IEEE'.
[25] IEEE: Ethically Aligned Design – A Vision for Prioratizing Human Well-being with Autonomous and Intelligent Systems, Version 2, 2017, 22–32. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf (Date of download: 25.06.2022.).

and intelligent systems; therefore, they are also relevant for intelligent cars. All in all, the guidelines must be considered during the design process.

3.3. The European Union

3.3.1. Ethical documents of AI

Recognising the potential of technology to boost the economy and competitiveness – e.g. social welfare, digital infrastructure, education and research –the European Union aims to lead the way in digitalisation through various developments related to artificial intelligence. As explained in our previous study on the regulation of AI: *"There are two expectations for the regulation: it must serve both the citizens and technology, as an insufficient regulation can lead to various infringements of citizens' rights and can be an obstacle to future development."*[26]

As a result, the European Commission established a group of independent experts to study the ethical aspects of the technology in 2018. The high-level expert group published its research results, 'Ethics Guidelines for Trustworthy AI', in April 2019. In this document, the EU intended to create a human-centred AI, highlighting three components – lawfulness, robustness and ethics – that should be applied throughout the entire lifecycle of the system.[27] All elements are necessary for the goal of trustworthiness, which must work in harmony as a whole.[28] As in the definition of the three essential elements of trustworthy AI, the document refers to the possibility of a conflict between the principles but does not propose a solution to resolve the problem, highlighting that *"…the legislature or policy-makers may need to review the adequacy of existing law where these might be out of step with ethical principles."*[29]

In the context of trustworthy AI, which constitutes lawful, ethical and robust AI, the Ethics Guidelines do not examine the issue of lawfulness in detail but rather focus on the latter two areas. Trustworthy AI is based on the fundamental rights – guaranteed by the European Charter and other international documents – including respect for human dignity, equality and solidarity, while also aiming to ensure that fundamental rights are respected. The four ethical principles derived from fundamental rights are also the basis of trustworthy AI and therefore need to be taken into account in the design process and application:

1. Based on the *principle of respect for human autonomy*, technology should not unduly restrict individuals' right to self-determination and freedom but instead support individuals.[30]
2. According to the *principle of prevention of harm, "AI systems should neither cause nor exacerbate harm or otherwise adversely affect human beings. This entails the protection of human dignity as well as mental and physical integrity."*[31]
3. *The principle of fairness* is interpreted quite broadly in the EU document, including material, procedural and social fairness.

[26] Ibolya STEFÁN: A mesterséges intelligencia jogi szabályozásának egyes kérdései, *Miskolci Jogi Szemle* 2020/3, 185.
[27] Independent High-level Expert Group on Artificial Intelligence: *Ethics Guidelines for Trustworthy AI, Brussels,* 8 April 2019, 5. https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf (Date of download: 10.06.2022.).
[28] Independent High-level Expert Group on Artificial Intelligence: op. cit., 5.
[29] Independent High-level Expert Group on Artificial Intelligence: op. cit., 5, 11.
[30] Independent High-level Expert Group on Artificial Intelligence: op. cit., 10–12.
[31] Independent High-level Expert Group on Artificial Intelligence: op. cit., 12.

4. The principle of explicability is intended to ensure trust and transparency in the operation of technology to individuals. [32]

A cardinal element of the Ethical Guidelines is the requirement for implementing trustworthy AI, which must be met throughout the entire 'lifecycle' of AI. We must note that we only list the requirements but do not examine them in detail.

1. Human agency and oversight,
2. Technical robustness and safety,
3. Privacy and data governance,
4. Transparency,
5. Diversity, non-discrimination and fairness,
6. Societal and environmental wellbeing,
7. Accountability.

Concerning our topic, 'embedded ethics' should also be mentioned, as it is a crucial element of the Ethics Guidelines because it is a technical method for implementing trustworthy AI. The essence of embedded value methods is that the standards to be followed are embedded in the technology. A link is established between the decisions taken by the AI and the principles to be followed, is in order to avoid negative impacts of the decision of the technology.

Later in 2020, the European Commission published '*White paper on Artificial Intelligence – A European approach to excellence and trust',* which also envisages building a trustworthy ecosystem.[33] Both documents aim to develop a common EU approach and provide guidance to the Member States in making their AI strategies. Effective regulation is essential to unleash the potential of technology and protect citizens' rights.

In the same year, the European Parliament issued its resolution *'Framework of ethical aspects of artificial intelligence, robotics and related technologies'*. The document has several essential findings; regarding the subject of our paper, recommendations on transport are worth mentioning. The resolution highlights the benefits of the technology, such as safer transportation and minimalisation of accidents (e.g. better reaction time). Recommends an up-to-date regulatory framework on the use of emerging technologies and an ethical framework establishing trustworthy AI. It also envisages the modernisation of infrastructure for the use of novelty in transportation.[34]

3.3.2. Ethics of connected and automated vehicles

The significance of ethics concerning intelligent cars has also been recognised quite early on. The result of the topic examination embodied in *'Ethics of connected and automated vehicles – Recommendations on road safety, privacy, fairness, explainability and responsibility'*[35]. The

[32] Cf. Independent High-level Expert Group on Artificial Intelligence: op. cit., 12–13.
[33] White paper on Artificial Intelligence – A European approach to excellence and trust, European Commission, Brussels, 19.2.2020, COM(2020) 65 final. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (Date of download: 10.06.2022.).
[34] Framework of ethical aspects of artificial intelligence, robotics and related technologies – European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL)), P9_TA(2020)0275, Brussels, 20.10.2020. https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_EN.pdf (Date of download: 13.09.2022.).
[35] Hereinafter referred to as 'Ethical Recommendation'.

document focuses on twenty recommendations and their interpretation regarding the ethical regulation of intelligent vehicles.

According to the recommendation, the following ethical guidelines are significant – and should be applied – regarding the design and implementation of connected automated vehicles (hereinafter referred to as 'CAVs'):

1. *Non-maleficence*: the application of CAVs cannot endanger living beings and the planet's integrity, nor increase the risk of harm for the participants of transportation in comparison to manual driving.
2. *Beneficence*: automated vehicles should improve and contribute to the well-being of individuals during the design and application phases.
3. *Dignity*: it gives the essence of human rights. Regarding CAVs, protecting human dignity demands preventing the violation of human rights in the service of other societal goods.
4. *Personal autonomy*: In the case of automated vehicles, it is necessary to protect the ability of people to make decisions and set their standards and goals. Among many others, it requires the protection of users against unreasonable restrictions, as well as hidden and aggressive marketing.
5. *Responsibility*: considering CAVs, it is vital to form proper moral and legal standards of responsibility to promote liability (legal and moral) for individuals and stakeholders. [36]
6. *Justice*: *"In relation to CAVs, that would mean, among other things, that CAVs should provide equality of access to mobility for all and should be calibrated by developers to reduce disparities in exposure to harm between categories of road users."*[37]
7. *Solidarity*: it means pro-social actions and practices. In the case of automated vehicles, it involves data-sharing regarding fatalities and injuries.[38]
8. *Inclusive deliberation: "The above principles cannot be applied with a mechanical top-down procedure. They need to be specified, discussed and redefined in-context. Inclusive deliberation ensures that perspectives from all societal groups can be heard, and no one is disregarded. Moreover, tensions between these principles may arise in specific applications. That is why the design and development of CAV systems should be supportive of and resulting from inclusive deliberation processes involving relevant stakeholders and the wider public."*[39]

The document also highlights a few recommendations considering responsibility, which is also essential in a legal sense. In accordance with the Ethical Recommendation there are five types of responsibility, such as: virtue; accountability; obligation; culpability and liability. In the following, we mention these briefly, considering ethics.

Recommendation 17 *"Promote a culture of responsibility with respect to the obligations associated with CAVs."*[40] This recommendation means responsibility as virtue and promotes

[36] Directorate-General for Research and Innovation: Ethics of Connected and Automated Vehicles, Luxembourg, 2020, 21–22. https://op.europa.eu/en/publication-detail/-/publication/89624e2c-f98c-11ea-b44f-01aa75ed71a1/language-en (Date of download: 28.06.2022.).
[37] Directorate-General for Research and Innovation: op. cit., 22.
[38] Directorate-General for Research and Innovation: op. cit., 22.
[39] Directorate-General for Research and Innovation: op. cit., 23.
[40] Directorate-General for Research and Innovation: op. cit., 10.

the development of moral agency; the absence of culture of responsibility in society and among individuals may result in the loss of empathy and autonomy.[41]

Recommendation 18 "*Ensure accountability for the behaviour of CAVs (duty to explain).*"[42] In this respect, the Ethical Recommendation refers to responsibility as accountability, the duty to explain, which functions as reinforcement and moral empowerment that holds the community together in a moral sense. Lack of accountability may lead to trust and public legitimacy issues.[43]

Recommendation 16 "*Identify the obligations of different agents involved in CAVs.*"[44] The document considers responsibility differently, as an obligation. According to this perspective, the aim is to avoid or reduce the amount of future damage; not fulfilling these criteria risks moral-ethical norms and trust problems.[45]

Recommendation 19 *"Promote a fair system of attribution of moral and legal culpability for the behaviour of CAVs."*[46] Responsibility also means legal culpability in the Ethical Recommendation; it is important to punish someone for an accident or error of the CAV to assure rehabilitation, deterrence and retribution,[47] which is essential both on the level of society and individual.

Recommendation 20 *"Create fair and effective mechanisms for granting compensation to victims of crashes involving CAVs."*[48] Last but not least, liability should be mentioned regarding responsibility. Liability aims to compensate victims; the lack of this would weaken the value of corrective justice. [49]

It should be noted that ethical-moral responsibility and legal (sectoral) responsibility – liability – are distinct areas of the ethical-moral normative environment that arise throughout intelligent cars' lifecycle. However, the law is also intended and capable of channelling moral rules, subject to certain limits. This perspective can be found in the German regulation of intelligent vehicles, which rules are established in light of the recommendations of Ethics Commission.[50]

4. Trolley problem – an ethical issue raised by the appearance of intelligent cars

As new technologies have appeared, ethical concerns about their use have emerged.[51] Regarding intelligent cars, one of the ethical issues is the so-called *'Trolley problem'*, which

[41] Directorate-General for Research and Innovation: op. cit., 54.
[42] Directorate-General for Research and Innovation: op. cit., 10.
[43] Directorate-General for Research and Innovation: op. cit., 54.
[44] Directorate-General for Research and Innovation: op. cit., 9.
[45] Directorate-General for Research and Innovation: op. cit., 54.
[46] Directorate-General for Research and Innovation: op. cit., 10.
[47] Directorate-General for Research and Innovation: op. cit., 54.
[48] Directorate-General for Research and Innovation: op. cit., 11.
[49] Directorate-General for Research and Innovation: op. cit., 54.
[50] See more, Federal Ministry of Transport and Digital Infrastructure: Ethics Commission – Automated and Connected Driving (Report), 2017. https://bmdv.bund.de/SharedDocs/EN/publications/report-ethics-commission-automated-and-connected-driving.pdf?__blob=publicationFile (Date of download: 25.06.2022.)
[51] Cf. Gergely László SZŐKE – Balázs HOHMANN: A „Trolley problem" gondolatkísérlet vonatkozásai az autonóm járművek terén – Mi a zsinórmérce a gépi döntéshozatal során? in: *Az autonóm járművek és intelligens rendszerek jogi vonatkozásai* (eds.: Judit LÉVAYNÉ FAZEKAS – Gábor KECSKÉS). Universitas-Győr Nonprofit Kft., Győr, 2020, 303–326.

refers to the forced decision-making of a person in a tense situation. The *'Trolley problem'* can be discovered in the study of *Philippa Foot*,[52] titled 'The Problem of Abortion and the Doctrine of the Double Effect'.[53] However, the term *'Trolley problem'* appeared first in the study of *Judith Jarvis Thompson*, called 'Killing, Letting Die and the Trolley Problem', in which she addresses the difficulty – which is the essence of the dilemma – as follows: *"Edward is the driver of a trolley, whose brakes have just failed. On the track ahead of him, are five people; the banks are so steep that they will not be able to get off the track in time. The track has a spur leading off to the right, and Edward can turn the trolley onto it. Unfortunately there is one person on the right-hand track. Edward can turn the trolley, killing the one; or he can refrain from turning the trolley, killing the five."*[54]

In this situation, the questions are: Which is the better option? Which act causes less harm? Which option is more acceptable and ethical from the perspective of the person and the community? Unfortunately, these questions cannot be answered because every option is wrong and unacceptable as someone – even if it is one person – will die. This controversy causes the dilemma which often appears regarding intelligent vehicles.

To help to ease the ethical dilemmas of the decision-making of intelligent or self-driving cars, a group of researchers from the Massachusetts Institute of Technology established an online platform called *'Moral Machine'*. The research took place globally as the *"…platform gathered 40 million decisions in ten languages from millions of people in 233 countries and territories."*[55] The platform aims to assess the social expectations of self-driving cars through accident scenarios, which the Moral Machine generates. In the test, the individuals can choose what should the intelligent car do. In an unavoidable accident scenario, the participant must choose from two options, which result in a lethal accident for some while saving others' lives. The survey focuses on several factors, such as species (humans – animals), the people involved in traffic (passengers – pedestrians), the number of persons (more lives – fewer lives), genders (men – women), age groups (old – young people), the health of individuals (fit – overweight), social statuses (doctor – criminal). There is a broad unanimity in the responses regarding three preferences: choosing humans over animals, young over elderly and minimalising the loss of lives in the accident. However, in other questions, the research results vary depending on geographic areas, highlighting the cultural differences. According to the established cultural

[52] The Trolley problem' stems from the so-called 'Double Effect', which was mentioned at first in 'Summa Theologiae', the work of Thomas Aquinas, considering the lawfulness of killing a person as self-defence. *"Accordingly the act of self-defense may have two effects, one is the saving of one's life, the other is the slaying of the aggressor. Therefore this act, since one's intention is to save one's own life, is not unlawful, seeing that it is natural to everything to keep itself in "being," as far as possible….it is lawful to repel force by force, provided one does not exceed the limits of a blameless defense."* St. Thomas AQUINAS: *Summa Theologiae*, 1485, Second Part, Question 64, Article 7. https://www.newadvent.org/summa/3064.htm#article7 (Date of download: 10.07.2022.).

[53] *"To make the parallel as close as possible it may rather be supposed that he is the driver of a runaway tram which he can only steer from one narrow track on to another; five men are working on one track and one man on the other; anyone on the track he enters is bound to be killed. In the case of the riots the mob have five hostages, so that in both the exchange is supposed to be one man's life for the lives of five."* Philippa FOOT: The Problem of Abortion and the Doctrine of the Double Effect, *Oxford Review* 1967/5, 2.

[54] Judit Jarvis THOMPSON: Killing, Letting Die and the Trolley Problem, *The Monist* 1976/2, 206.

[55] Edmond AWAD – Sohan DSOUZA – Richard KIM – Jonathan SCHULZ – Joseph HENRICH – Azim SHARIFF – Jean-François BONNEFON – Iyad RAHWAN: The Moral Machine Experiment, *Nature* 563(7729) (2018), 59.

clusters, there are different perspectives on the age groups and social statuses (e.g. spare for younger or higher status people is less relevant in the Eastern cluster than in the South).[56]

5. Closing remarks

In this paper, we have studied the technological background of the technology to understand the novelties' operation and highlight its importance. Moreover, we have reviewed the relevant provisions of the most important international documents regarding the ethics of intelligent cars and briefly described one of the most important ethical dilemmas.

Several documents' catalogues of values and principles show that the goal is to design a human-centric AI. Therefore, the recommendations and guidelines focus on the universal, equal human rights which must be implemented into the technology and protected in every lifecycle – from design to operation – of intelligent vehicles. Not only human rights are essential, but other ethical values – e.g. justice and responsibility – which are respected, accepted and applied by the entire society, decisive in human relations. It should be noted that ethical values and principles have a significant effect on legal relations, specifically on civil and criminal liability. Legal rules also involve ethical-moral values to a specific limit, for example, the principles of each code which are determining, as they have a deeper meaning and affect the entire legal system as they must be considered in legal relations and procedures.

The documents mentioned earlier show that the 'creation' of new technologies and their legislation require the collaboration of different disciplines – engineers, lawyers, and philosophers – as they interact throughout the process. Therefore, programmers need to consider specific social values, and so do lawyers need to assess the specifics of the technology when drafting legislation to find the proper combination.

[56] AWAD – DSOUZA – KIM – SCHULZ – HENRICH – SHARIFF – BONNEFON – RAHWAN: op. cit., 60–61.