

CHARLES UNIVERSITY

Faculty of Law

Martin Vlasta

**Methodological framework for European regulation of (artificial)
intelligence and its limits**

Master's thesis

Master's thesis supervisor: prof. JUDr. PhDr. Michal Tomášek, DrSc.

Department of European Law

Date of completion (manuscript closure): [04. 06. 2023]

UNIVERZITA KARLOVA

Právnická fakulta

Martin Vlasta

**Metodologický rámec pro Evropskou regulaci (umělé) inteligence
a její limity**

Diplomová práce

Vedoucí diplomové práce: prof. JUDr. PhDr. Michal Tomášek, DrSc

Katedra: Evropského práva

Datum vypracování práce (uzavření rukopisu): [04. 06. 2023]

Prohlašuji, že jsem předkládanou diplomovou práci vypracoval/a samostatně, že všechny použité zdroje byly řádně uvedeny a že práce nebyla využita k získání jiného nebo stejného titulu.

Dále prohlašuji, že vlastní text této práce včetně poznámek pod čarou má 146 558 znaků včetně mezer.

Martin Vlasta

V Praze dne 4. 6. 2023

I declare that this master's thesis is a result of my independent work and that all sources used have been duly quoted; and that this master's thesis was not used to obtain other or the same degree.

I further declare that the body of this master's thesis, including footnotes, has 146 558 characters.

Martin Vlasta

In Prague on 04 July 2023

Poděkování

Rád bych poděkoval panu prof. JUDr. PhDr. Michalu Tomáškoví, DrSc., za odborné rady a strategické vedení při sepisování textu práce. Dále bych rád poděkoval JUDr. Jaroslavovi Denemarkovi za praktické nasměrování v oblasti existující judikatury a poznatky v oblastech nové evropské legislativy.

Acknowledgements

I would like to thank Prof. JUDr. PhDr. Michal Tomášek, DrSc., for his expert advice and strategic guidance during the writing of the thesis. I would also like to thank JUDr. Jaroslav Denemark for his practical guidance in the field of existing case law and knowledge in the areas of new European legislation.

Contents

- Contents.....5
- 1. Introduction.....7
 - 1.1. Goals of research7
 - 1.2. Methods used.....7
 - 1.3. Context of the topic of the work8
- 2. Why regulate artificial intelligence?9
 - 2.1. The mantle of responsibility and decision-making.....10
 - 2.2. Examples of transfer of decision-making ability from a human to a machine.....12
- 3. Legislative roadmap.....13
 - 3.1. Actors involved in the legislative process13
 - 3.2. Change in the narrative from opportunity to threat14
 - 3.3. Guiding values analysis and framing regulatory intent16
 - 3.4. Summary of protoregulatory history18
 - 3.5. Values extracted from AI proto-regulation and their mirror image in legislation.....22
 - 3.6. Calls for decisive action and the European legislative process24
 - 3.7. The structure of the regulation and comparison of positions for legislative trilogue.....25
 - 3.8. Methodology of regulation (Law)31
 - 3.9. Introduction to Regulation Models for AI.....32
- 4. What is (should be) Artificial intelligence from regulatory perspective?.....34
 - 4.1. Lessons learned from the human condition34
 - 4.2. AI according to AI scientists34
 - 4.3. AI according to Policy makers36
 - 4.1. Analysis of the functional features of AI and legislative recommendations.....37
 - 4.1.1. Epistemology (Perception and acquisition of data and knowledge)38
 - 4.1.2. Cognition (Information and data processing and decision making)40
 - 4.1.3. Agency (ability to affect the world)44
 - 4.1.4. Autonomy (adaptability, self-governance, self-actualization, self-repair).....45
 - 4.1.5. Purpose (reasons and intent, goals and desires)46
- 5. Arising new AI-specific rights and duties.....49
 - 5.1. Potential AI-specific rights51
 - 5.1.1. Right to erasure own data from a data model.....51
 - 5.1.1. Right to corrective action in a data model.....52
 - 5.1.1. Right to explanation52
 - 5.1.2. Right to disclosure.....53

5.1.3. Right to redress and human review	54
6. Conclusion	56
7. List of abbreviations	60
8. Sources	61
8.1. Legislation	61
8.2. Proposed legislation.....	62
8.3. Case Law	63
8.4. Literature sources and academic papers	63
8.5. Official publications and advisory documents	65
8.6. Other sources	68
9. Abstrakt v českém jazyce – Metodologický rámec pro Evropskou regulaci (umělé) inteligence a její limity	70
10. Abstract in English language - Methodological framework for European regulation of (artificial) intelligence and its limits	71

Rozsah vlastního textu diplomové práce (úvod, jednotlivé části a závěr) včetně poznámek pod čarou je nejméně 108 000 znaků včetně mezer vlastního textu.

1. Introduction

1.1. Goals of research

The goal of this paper is to evaluate methodological approaches to regulation of artificial intelligence, specifically in the instance of the upcoming European legislation, namely the Regulation on Artificial Intelligence¹, introduced by the European Commission into the trilogue in 2023².

First, the paper will answer the question what values are driving the legislative intent and whether it will be in effect in time.

Second, what method of regulation was used for the Regulation on Artificial Intelligence and what type of new responsibilities and law enforcement mechanisms it aims to introduce.

Finally, the paper will also explore possible new future AI-related rights and duties that either arise from new or existing legislation (*de lege lata*) or that have been omitted but should arguably have been there to provide more clarity (*de lege ferenda*).

1.2. Methods used

To answer the first series of questions, throughout the paper, a diverse set of sources relating to normative demands is explored and analyzed. Ranging from international bodies such as UN³, OECD and WEF⁴ to academic research groups, a notable highlight would be the Berkman Klein Center for Internet & Society at Harvard University⁵ and finally the research bodies at the European Union, specifically the Joint Research Centre (JRC)⁶, the European Commission's science and knowledge service. All of these research initiatives are laying the groundwork for the future legislative act that has taken form within the proposed regulation.

¹ EC – Proposal for a Regulation of the European Parliament and of the Council on Artificial Intelligence. Eur-lex.europa.eu. [online]. [cit. 2021-04-21]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>

² EP – Press Release AI Act: a step closer to the first rules on Artificial Intelligence. Europarl.europa.eu. [online]. [cit. 2023-05-11]. Available from: <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>

³ UN – Report: The Age of Digital Interdependence. Un.org. [online]. [cit. 2019-June]. Available from: <https://www.un.org/en/pdfs/HLP%20on%20Digital%20Cooperation%20Report%20Executive%20Summary%20-%20ENG.pdf>

⁴ WEF – White paper: How to Prevent Discriminatory Outcomes in Machine Learning. World Economic Forum – Global Future Council on Human Rights 2016-2018. Weforum.org [online]. [cit 2018-03-01]. Available from: https://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf

⁵ Harvard – Timeline of Principled Artificial Intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center for Internet & Society at Harvard University. cyber.harvard.edu [online]. [cit. 2020-01-15]. Available from: <https://cyber.harvard.edu/publication/2020/principled-ai>

⁶ EC – AI Watch. Defining Artificial Intelligence 2.0. publications.jrc.europa.eu [online]. [cit. 2021-10-29]. Available from: <https://publications.jrc.ec.europa.eu/repository/handle/JRC126426>

To answer the second question, what is the nature and method of the regulation, a framework laid down by N. Petit and J. Cooman⁷ is utilised and through prism of his established methodology the regulation is assessed.

Finally, to answer the last question, what possible new AI rights and duties might we expect, current legislation that is in effect is explored, as well as its practical application in the real world with provided examples of case law by Court of Justice of the EU within its case law.

1.3. Context of the topic of the work

The volume of data produced in the world is growing rapidly⁸. Exponential growth of data and data processing capability brings about many opportunities but also challenges for the regulator. It is of vital importance to join forces between multiple different disciplines and combine collective knowledge to create supportive and successful environment for further innovation.

There are real fears that overregulation might lead to preemptively neutering any meaningful advancement in essential fields, as is often criticised in many fields of industry such as healthcare⁹ ¹⁰, finance¹¹. The authors suggest that overburdening the industry results in an effective brain-drain and missed opportunity of untapped potential benefits that can be garnered from effective AI application usage across many fields.

The game theory and market dynamics are also in effect. From that standpoint, it is reasonable to assume that some form of forum-shopping¹² might take place in regards to development of AI applications, and that data scientists and companies will flock to a jurisdiction that has least barriers and lowest cost of compliance and operate from there.

At the end of the day, I believe it is right that the European Union continues with its ambitious plans and leads by example to inspire other jurisdiction to adapt the same mechanisms and values, which will lead to a more secure, connected and cohesive world.

⁷ PETIT, N., DE COOMAN, J. – Models of Law and Regulation for AI. Robert Schuman Centre for Advanced Studies Research Paper No. RSCAS 2020/63, EUI Department of Law Research Paper papers.ssrn.com. [online]. [cit. 2020-10-08]. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3706771

⁸ European Parliament, Artificial intelligence: threats and opportunities europarl.europa.eu (online). [cit. 2020-05-04]. Available from: <https://www.europarl.europa.eu/news/en/headlines/society/20200918STO87404/artificial-intelligence-threats-and-opportunities>

⁹ GREENBAUM, D. (2018). Avoiding Overregulation in the Medical Internet of Things. In I. Cohen, H. Lynch, E. Vayena, & U. Gasser (Eds.), *Big Data, Health Law, and Bioethics* (pp. 129-141). Cambridge: Cambridge University Press. doi:10.1017/9781108147972.013

¹⁰ SHEN, N. AI Regulation in Health Care: How Washington State Can Conquer the New Territory of AI Regulation, 13 *Seattle J. TECH. ENV't & INNOVATION L.* 1 (2023).

¹¹ KING, S., Why Bigger Is Not Always Better: Dodd-Frank & Its Impact on Small Banks & Businesses, 36 *J. C.R. & ECON. DEV.* 78 (2022).

¹² HARRISON, J., & WOODS, L. (2007). Jurisdiction, forum shopping and the 'race to the bottom'. In *European Broadcasting Law and Policy* (Cambridge Studies in European Law and Policy, pp. 173-193). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511495298.010

Long term benefits of AI usage are so profound they are basically unrejectable. Our society must get there one way or another, but it will require a change of thinking, a trust that goes beyond an individual, beyond a family, beyond a country. It must be trust towards an idea, common culture and values embedded in code. Such trust can be grounded in unquestionable belief (religion), absolute obedience (fascism/totalitarianism), or understanding and knowledge (democracy). It is the ambition of the European Union to promote democratic values not in spite of, but rather empowered by the AI applications of the future¹³.

If these systems are developed behind closed doors, obfuscated, difficult to understand, smeared by corruption and overpriced public procurement contracts, vendor lock-ins, private negligence and sole profit extraction driven intents, then any social contract built upon such bad faith cannot stand to prove itself as sustainable and will eventually succumb to market failure.

We should all strive to learn about how these new AI systems work, about its internal checks and ballances, about how the data is stored and used, about how it affects our lives and lives of others, because only in knowledge we can confidently opt in and take part in such society.

2. Why regulate artificial intelligence?

It is established that law is a practical, functional and goal oriented system¹⁴. Regulating AI seems to be a no-brainer for most. But what are we actually trying to achieve? What drives our need for AI regulation?

We have reached a technological point, where AI systems today (whatever that means will be explored in later chapter) have the ability, capacity to inflict harm, including loss of life, damage to property, societal damage, aggravate division and enhance existing latent discrimination on system-wide scale¹⁵, completely unchecked, unsupervised, unquestioned, without means of remedy. Many of these critical risks have been pointed out in multitude of studies.¹⁶ Humans sense that they are powerless in face of a potentially ever perfect machine. We use law to regulate excessive power. We have delegated our decision making, of complex matters (i.e. power) to

¹³ European Parliament, Artificial intelligence: threats and opportunities europarl.europa.eu (online). [cit. 2020-05-04]. Available from: <https://www.europarl.europa.eu/news/en/headlines/society/20200918STO87404/artificial-intelligence-threats-and-opportunities>

¹⁴ GERLOCH, Aleš. Teorie práva. 7. vyd. Plzeň: Aleš Čeněk s.r.o., 2017. ISBN 978-80-7380-652-1. s

¹⁵ WEF – White paper: How to Prevent Discriminatory Outcomes in Machine Learning. World Economic Forum – Global Future Council on Human Rights 2016-2018. Weforum.org [online]. [cit 2018-03-01]. Available from: https://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf

¹⁶ UN – Report: The Age of Digital Interdependence. Un.org. [online]. [cit. 2019-June]. Available from: <https://www.un.org/en/pdfs/HLP%20on%20Digital%20Cooperation%20Report%20Executive%20Summary%20-%20ENG.pdf>

automated rule based processes (AI), and there is no feasible method to doublecheck or review all of the decisions done autonomously.

Due to the number of decisions required for us to get through our lives, while having only limited human energy and attention span, to increase productivity and wealth, we filtered out certain rudimentary decisions we do not want to process anymore and introduced a number of rule based systems.

Much like when a manager of a company decides to hire an assistant, and slowly transfers his job - task by task - to said assistant, and in the end, the director kept for himself the most important decisions, but in the process forgot what those rules even are, what does the company even do, and he grew so dependent on said assistant that he basically became his slave.

Law is one of said rule-based systems as well. When conflict occurs, we don't have the time to think about ethics and abstract rule ad hoc on the spot, discover all of the relevant facts, universally reach out to everyone to ask their opinion and deliver justice. We deferred our decision making of what is just and what is wrong to a system of rules on paper, professional judicial system, and bureaucratic administration to attain the most feasible accurate cost-efficient decision-making system possible. But even that came full circle. Law became itself too complicated to be accessible.

The common theme is transferring decision-making. The entire history of mankind can be painted as a narrative between a sovereign and those that are governed. On individual level it can be a child that wishes to not be governed by his father anymore. On a collective level a group of people that shares some common trait (i.e. ethic, cultural, wealth) wishes to not be governed by another group of opposite trait. On a global level, a free city wishes not to be governed by central authority in far away land.

2.1. The mantle of responsibility and decision-making

Throughout history vast majority of society was mostly not free to do what they wished. Law can be used as a means of oppression and as a means of ending or mitigating oppression¹⁷. There was a time, when slavery was legal and an important part of nations' economies¹⁸. Eventually through different political stages we have arrived to the current liberal world-view with individual responsibility and freedoms as the centre of our legal system, as is evidenced by

¹⁷ MAY L. (2019). Slavery and Democracy. In *Ancient Legal Thought: Equity, Justice, and Humaneness From Hammurabi and the Pharaohs to Justinian and the Talmud* (pp. 290-302). Cambridge: Cambridge University Press.

¹⁸ SWANSON E., & BREWER STEWART, J. (2018). Defending Slavery, Denying Slavery: Rhetorical Strategies of the Contemporary Sex Worker Rights Movement in Historical Context. In E. Swanson & J. Stewart (Eds.), *Human Bondage and Abolition: New Histories of Past and Present Slaverys* (Slavery since Emancipation, pp. 262-293). Cambridge: Cambridge University Press. doi:10.1017/9781316890790.011

multitudes of charters¹⁹ and constitutions written up thereafter. The idea was to take back control that was forcefully withheld from the emperors or slaves' owners back then.

The basic premise in the current legal system is that there is a stark difference between a subject and object²⁰. Subjects are assigned rights and duties and the ultimate responsibility for their actions or inactions, be it criminal²¹, administrative²² or civil²³. As the legal system developed, it introduced legal persons²⁴, and assigned responsibilities to them also²⁵.

In modern age, these responsible actors transferred the mantle of responsibility and decision-making to rule-based systems, which in view of the law are mere objects, yet they are making effective decisions and responsible actors rely on these decisions unquestionably.

Once again we are coming full circle, where we feel like we are losing our sense of freedom in light of ever increasing governing pressure from rule-based systems. As long as we can identify with said rule-based system, we do not view it as oppression, but rather as self-legislation, which is welcomed.²⁶ This is why Kant's ideas were so ahead of his time in my opinion. This is also why democracy based systems are functional, because the recipients of rules are – although often times through a very convoluted process – also creators of said rules or are involved in the process through consultation.

Legal rules can more or less be understood by every recipient, because the decision-making process can be comprehended and traced back to clear set of „if > then“ rules that have been agreed to prior. Automation of decision-making allows nesting of multiple „if > then“ rules together to create very complex decision-making trees.

As long as we can understand the decision-making process, we are more likely to accept the decision. In the case where we are the agents of decision-making, we are more likely to accept our own conclusion because we believe our process was flawless. In the case where someone else is the agent of decision-making (such as asking a calculator what the result is of an equation) then I am not involved in the decision-making itself, i only know input (i.e. question) and output (answer, or decision), and i need to trust that the process is as flawless as myself. If it is not flawless, then I would rather do it myself, to make sure it is flawless.

¹⁹ Preamble to the Universal Declaration of Human Rights

²⁰ GERLOCH, Aleš. *Teorie práva*. 7. vyd. Plzeň: Aleš Čeněk s.r.o., 2017. ISBN 978-80-7380-652-1. s. 156

²¹ Section 12 et seq. of Act No. 40/2009 Coll., the Criminal Code

²² Section 5 et seq. of Act No. 250/2016 Coll., the Act on Liability for Offences and Proceedings Thereon

²³ Section 2894 et seq. of Act No. 89/2012 Coll., the Civil Code

²⁴ Section 18 et seq. of Act No. 89/2012 Coll., the Civil Code

²⁵ Section 7 et seq. of Act No. 418/2011 Coll., the Act on Criminal Liability of Legal Persons and Proceedings Against Them

²⁶ PATRICK, K. "Self-legislation in Kant's Moral Philosophy," *Archiv für Geschichte der Philosophie*, 2004, 86(3): 257–306.

AI based systems are employed in areas, where we do not have human capacity to do decision-making, because it is too expensive or there is high probability of human error. Decisions done by AI are accepted only when they are flawless, because otherwise we would have done them ourselves. The problem is, we have delegated the power of decision-making to systems, which are not flawless, they can and do cause harm and there are very few options how to correct it.

This is the triangle of economic decision-making. I want the best thing, I want it now, and for free. Best output, no transaction cost, least input. Every decision-making process has to go through this triangle. In case of AI, we have gone from a world where humans were the attainable best, but they are very expensive (but still the least expensive at the time) and took a long time (because there wasn't anything faster) to a world where machines are the best, the fastest and consume less input (energy or food) than humans. Problem is, when mistake is done by a human, we have systems to rectify that, we can even ask him or her, why did he act that way and adjust for next time. But what about an AI?

2.2. Examples of transfer of decision-making ability from a human to a machine

The reason we sense the urge to regulate AI is not because of the AI or machine itself, but because we need to put limits and rules on decision-making process itself. Under what conditions can a decision be deferred? How do we shape the decision making structures of shared or augmented decision-making where man and machine collaborate as cognitive hybrids? How to behave when a wrong decision is made? What should be the governing principles behind decision-making?

To really achieve what we all sense but can't quite grasp, is not to regulate artificial intelligence, but intelligence itself instead. This revelation has absolutely shattering consequences. We are about to create a generalised method of audit that is to be applied to all decision-making processes where automated autonomy is present, such as with AI based systems, but AI based systems are only abstraction of problems previously decided by humans. It is not far fetched to realise that what we will inevitably do, is that all decision-making is, task by task, transferred from mankind to the machine, and all decision-making ever will be audited through this future framework. Let's take a step back and think about how this might affect crucial legal processes and moments:

- Administrative discretion utilised by an official in administrative proceedings²⁷,
- Judgement of a judge or a jury in litigation proceedings²⁸

²⁷ SKULOVÁ, S. Správní uvážení. Brno: Masaryková univerzita Brno., 2003. ISBN 80-210-3237-5

²⁸ WINTEROVÁ, A., MACKOVÁ, A. a kol. Civilní právo procesní. První část - Řízení nalézací. Vysokoškolská učebnice. 8. nezměněné vydání. Praha: Leges, 2015.

- Policy-making of a legislator in legislative process
- Due diligence by a CEO of a private company in managerial decisions²⁹
- Deciding who to vote for during elections
- Deciding who to believe in terms of the news we consume

This list could be endless, and that is terrifying. How to decide when the decision-making agent is uncertain but we require a decision anyway? There are situations where any decision no matter how bad it is, is still better than hesitation and inaction.

One would get the impression that Regulation of AI would only affect AI, but through this realisation we see that it is by far one of the most fundamental pieces of laws that strikes at the very core of what makes us human, and that is personhood and rationale, intelligence, the capacity to think.

The way we are defining the rules for AI right now, leads to defining rules for „I“, for any intelligence system on Earth and beyond. We are about to enter a world, where „thought“ itself is to be regulated.

3. Legislative roadmap

3.1. Actors involved in the legislative process

There is a lot at stake here with the regulation of AI systems. There is already a body of legislation that deals with subsets of problems that relate to AI applications, such as aforementioned GDPR³⁰, that deals with processing of personal data, NIS2³¹ that lays down harmonised rules for cybersecurity measures, DGA³² introduced a series of new obligations for public bodies to share data, DSA³³ & DMA³⁴ are brand new pieces of legislation that tackle the

²⁹ KRÁL, Richard. Koncepce a obsah péče řádného hospodáře. Univerzita Karlova. Právnická fakulta, 2016, 74 s.

³⁰ EP – Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). eur-lex.europa.eu [online]. [cit. 2023-02-05]. Available from:

<https://eur-lex.europa.eu/legal-content/EN-CS/TXT/?from=CS&uri=CELEX%3A32016R0679>

³¹ EP – Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive) (eur-lex.europa.eu [online]. [cit. 2023-03-24]. Available from: <https://eur-lex.europa.eu/eli/dir/2022/2555>

³² EP – Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) eur-lex.europa.eu [online]. [cit. 2023-02-05]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022R0868>

³³ EP – Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (eur-lex.europa.eu [online]. [cit. 2023-02-05]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065>

³⁴ EP – Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) eur-lex.europa.eu [online]. [cit. 2023-02-05]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R1925>

big tech and introduce additional layer of compliance. Some of the bigger principles are already set in stone and are in force. But with the draft of the regulation on AI looming on the horizon, though subtly delayed due to present circumstances (Covid-19, War in Ukraine, etc.) it will inevitably find its place back at the table.

The European Union's Digital Strategy is unwavering and if its overzealous commitments found in ESG legislation and sustainability legislation (Fit for 55)³⁵ are of any indication, we may expect the same vigour with the Digital Single Market³⁶

³⁷. All of these pieces of regulatory framework have in common their exportability aspect beyond the jurisdiction of the European Union. We are looking at a global civilisational initiative that will potentially lay down the foundations for the use of intelligence world-wide for centuries to come.

There are many groups with diverging opinions and agendas that wish to have a say in the future legislation for AI systems. Business and industry are preoccupied with ease of access and liberty to innovate, produce and introduce series of products that may generate revenue for them. Consumers are worried about their individual rights and protections. Many already sense the powerlessness when faced with huge corporations, which create products that the consumers are fully dependent on. An environment of dependency and one-sided dominance of corporate power is one that begets market failures and a lot of potential harm. Many representatives from Academia have a more principled approach and care about correctness in regards to their field.

Since AI application is by design a cross-field discipline, it requires coordination and discussion from many different experts both from universities of science and humanities. Policy makers are preoccupied with overall balance of interest groups, political feasibility of proposals, geopolitical implications, and predominantly its effects on law enforcement, economic feasibility, and potential new tax arrangements. Regulatory impact assessment³⁸ (RIA) must be given its due dilligance in this regard.

3.2. Change in the narrative from opportunity to threat

³⁵ EC – Fit for 55 - European Green Deal. consilium.europa.eu. [online]. [cit. 2023-06-04]. Available from: <https://www.consilium.europa.eu/en/policies/green-deal/fit-for-55-the-eu-plan-for-a-green-transition/>

³⁶ EP – Regulation (EU) 2021/694 of the European Parliament and of the Council of 29 April 2021 establishing the Digital Europe Programme and repealing Decision (EU) 2015/2240. eur-lex.europa.eu. [online]. [cit. 2021-05-11]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32021R0694>

³⁷ EP – European Declaration on Digital Rights and Principles for the Digital Decade eur-lex.europa.eu [online]. [cit. 2022-01-26]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A28%3AFIN>

³⁸ EC – Commission Staff Working Document Executive Summary Of The Impact Assessment Report Accompanying The Proposal For A Regulation of the European Parliament and of the Council, eur-lex.europa.eu . [online]. [cit. 2021-04-21]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021SC0085>

The regulatory intent was first declared³⁹ on April 25th in 2018 when Commission published a communication establishing its goals for future work and research that needed to be done to deliver a harmonised set of rules for AI. The communication, while not being a legally binding instrument⁴⁰, is nonetheless important for interpretation of studies that followed thereafter. It must be taken into the account the circumstances of the year 2018. World was relatively peaceful and would not hear of COVID-19 pandemic for a year still, nor the War on Ukraine and Energy Crisis. World affairs were preoccupied with a trade war between USA and China⁴¹, and field of AI was seen as one of the possible venues of competition. The text of the communication was quite positive and hopeful. It spoke of boosting EU's technological capacity, socio-economic modernisation of education, and previous achieved successes such as unmanned agricultural vehicles, robotic ortho-prostheses that restore mobility to amputees, alleviation of repetitive tasks for workers in car manufacturing plants.⁴² It welcomed the AI revolution and wanted to position Europe as the world leader in the field and attract new investment opportunities. This positive spin on AI is in stark contrast with the way narrative changed when the regulation was released in 2021. Suddenly AI was viewed as a threat. Subsequent communication by European Commission in 2021⁴³ speaks of very high risks to safety and fundamental rights. It frames the regulation not as a technology enabler, but rather an urgently needed measure due to being faced with unprecedented rapid technological development of AI.

³⁹ EC – Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Artificial Intelligence for Europe eur-lex.europa.eu. [online]. [cit. 2018-04-25]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN>

⁴⁰ SVOBODA, Pavel. Úvod do evropského práva. 5 vyd. Praha : C. H. Beck. 2013., str. 118

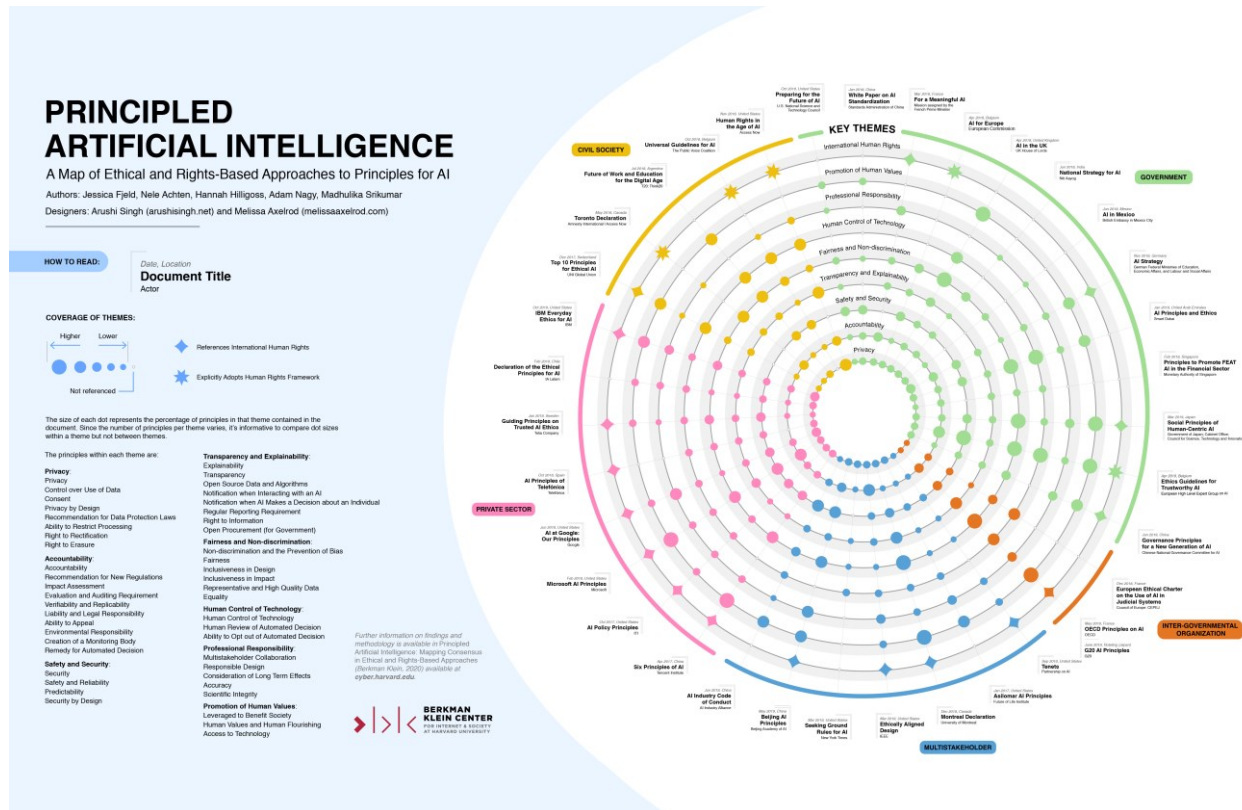
⁴¹ Carnegie Endowment for International Peace, The U.S.-China Trade War Has Become a Cold War. carnegieendowment.org [online]. [cit. 2021-09-16]. Available from: <https://carnegieendowment.org/2021/09/16/u.s.-china-trade-war-has-become-cold-war-pub-85352>

⁴² EC – Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Artificial Intelligence for Europe eur-lex.europa.eu. [online]. [cit. 2018-04-25]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN>

⁴³ EC – Communication on Fostering a European approach to Artificial Intelligence. digital-strategy.europa.eu. [online]. [cit. 2021-04-21]. Available from: <https://digital-strategy.ec.europa.eu/en/library/communication-fostering-european-approach-artificial-intelligence>

3.3. Guiding values analysis and framing regulatory intent

While the idea for a specific AI regulation is relatively new a lot of text has already been written that attempts to tackle the AI regulation topic from various angles. A helpful visualisation by Jessica Fjeld and Adam Nagy from Berkman Klein Centre at Harvard University published a helpful visualisation that captured what they deemed as most significant documents relation to AI regulation in the beginning of 2020.



Picture 1 – Visualization of documents written about Principled AI⁴⁴.

According to their report, it seems that the issues of privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination are securely at the forefront of the agenda for all authors irrespective of their background, whether it be private sector, civil society or government. It is relieving to see such a consensus, which significantly helps to frame the regulatory debate. Here is a list of other the often-mentioned phrases or keywords found in these documents:

- human-centric, inclusive, ethical,
- data protection, privacy, GDPR⁴⁵ compliant by design

⁴⁴ Harvard – Timeline of Principled Artificial Intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center for Internet & Society at Harvard University. cyber.harvard.edu [online]. [cit. 2020-01-15]. Available from: <https://cyber.harvard.edu/publication/2020/principled-ai>

⁴⁵ EP – Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the proceeding of personal data and on the free movement of such data, and

- cybersecurity, trust, reliability, robustness, safety
- fair, reasonable, non-discriminatory, diversity,
- international competitiveness, interoperability, accessible (libraries)
- user friendly, provide more choice and information
- lawful, compliant with other legislation by default
- environmental and societal well-being
- accountability, responsible usage, proper training and education
- human oversight, transparency and auditability, trustworthy

Member States are pointing at the current absence of a common European framework. The German Data Ethics Commission has called for a five-level risk-based system of regulation that would go from no regulation for the most innocuous AI systems to a complete ban for the most dangerous ones⁴⁶. Denmark has just launched the prototype of a Data Ethics Seal⁴⁷. Malta has introduced a voluntary certification system for AI⁴⁸.

All of these initiatives prompted a critique from the industry and in turn provided opportunity for research bodies to provide their first reaction on what the compliance model should look like. For example, in a paper AI Certification by P. Cihon⁴⁹ he argues that government certification programs should emphasize governance criteria of enduring value, so as to not age with the technology advancements.

If the EU fails to provide an EU-wide approach, there is a real risk of fragmentation in the internal market, which would undermine the objectives of trust, legal certainty and market uptake.⁵⁰ This notion was reaffirmed a year later in 2021 in the report on Coordinated Plan on AI

repealing Directive 95/46/EC (General Data Protection Regulation). eur-lex.europa.eu [online]. [cit. 2023-02-05]. Available from:

<https://eur-lex.europa.eu/legal-content/EN-CS/TXT/?from=CS&uri=CELEX%3A32016R0679>

⁴⁶ Data Ethics Commission of the Federal Government Federal Ministry of the Interior, Building and Community, Opinion of the Data Ethics Commission. Bmi.bund.de [online]. [cit. 2019-12-01] Available from: https://www.bmi.bund.de/SharedDocs/downloads/EN/themen/it-digital-policy/datenethikkommission-abschlussgutachten-lang.pdf?__blob=publicationFile&v=5

⁴⁷ Ministry of Industry, Business and Financial Affairs, New seal for IT-security and responsible data use is in its way. Eng.em.dk [online]. [cit. 2019-10-31] Available from: <https://eng.em.dk/news/2019/oktober/new-seal-for-it-security-and-responsible-data-use-is-in-its-way/>

⁴⁸ Parliamentary Secretariat for Financial Services, Digital Economy and Innovation, Towards Ethical and Trustworthy AI. Malta.ai [online]. [2019-08-03] Available from: https://malta.ai/wp-content/uploads/2019/08/Malta_Towards_Ethical_and_Trustworthy_AI.pdf

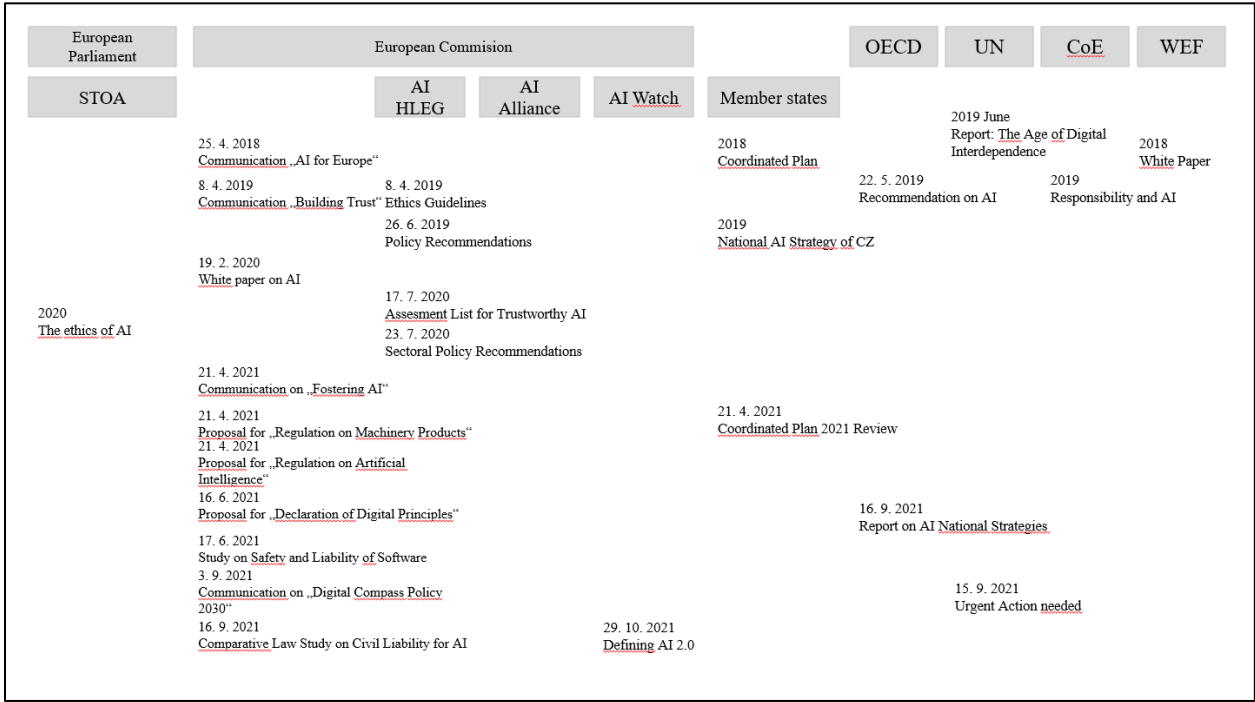
⁴⁹ CIHON, P., KLENALTENKAMP, M. J., SCHUETT, J., BAUM, S. D., AI Certification: Advancing Ethical Practice by Reducing Information Asymmetries, 2021, IEEE Transactions on Technology and Society, doi:10.1109/TTS.2021.3077595

⁵⁰ EC – White Paper on Artificial Intelligence: a European approach to excellence and trust. Ec.europa.eu. [online]. [cit. 2020-02-19]. Available from: https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

by the European Commission⁵¹. All of these documents are calling for action and express a sense of urgency.

3.4. Summary of protoregulatory history

Since 2020 the regulatory debate in the European Union has significantly matured through a series of EU-led initiatives, which produced a series of proto-regulatory accomplishments which laid down the foundations for the first ever draft of a European AI specific regulation. Below I introduce a follow up on this timeline with focus on European regulatory precursors that might be more impactful for future legislation:



Picture 2 – Timeline of proto-regulatory documents in the EU and affiliated partners

AI HLEG, stands for High-level expert group on artificial intelligence, appointed by the European Commission to provide advice on AI strategy, with task of writing up 4 key deliverables, which served as the foundation for further debate and analysis. The group in total had 51 members, with diverse backgrounds, such as representatives of academia or industry experts and representatives from the big tech. One of these members was Nicolas Petit, a Co-rapporteur from University of Liège, whose methodology for AI regulation is utilised in later chapter. During the first year of mandate, the AI HLEG body wrote up the Ethics Guidelines for Trustworthy AI.⁵² In

⁵¹ EC – Coordinated Plan on Artificial Intelligence 2021 Review. Digital-strategy.europa.eu. [online]. [cit. 2021-04-21]. Available from: <https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review>

⁵² AI HLEG – Ethics guidelines for trustworthy AI. Digital-strategy.ec.europa.eu [online]. [cit. 2019-04-08]. Available from: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

this document, the term „Trustworthy AI“ is coined. By their definition, it is a quality that fulfils following criteria⁵³:

- Lawful – respecting all applicable laws and regulation
- Ethical – representing ethical principles and values
- Robust – both from a technical perspective while taking into account its social environment

Moreover, it proposes series of 7 key requirements that AI systems should fulfil before they can be considered trustworthy⁵⁴.

- Human agency and oversight
- Technical Robustness and safety
- Privacy and data Governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and Environmental well-being
- Accountability

The first draft attracted more than 500 comments that were received through an open consultation. This further led to another deliverable titled Policy and Investment Recommendations for Trustworthy AI⁵⁵. In its 7th chapter it promotes the idea of adopting a risk-based approach to regulation and called for a systemic mapping and evaluation of all existing EU laws that are particularly relevant to AI systems in tandem, to avoid ‘silo-thinking’. Following areas of legal regulation were identified⁵⁶:

- Civil liability
- Criminal liability
- Consumer protection
- Data protection
- Non-discrimination
- Cybersecurity
- Competition and internal market

⁵³ Idem.

⁵⁴ Idem.

⁵⁵ AI HLEG – Policy and investment recommendations for trustworthy Artificial Intelligence. Digital-strategy.ec.europa.eu. [online]. [cit. 2019-06-26]. Available from: <https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence>

⁵⁶ Idem.

This was a very important turning point in the regulatory thought. Because up till now the regulation was viewed as a relatively minor piece of legislation that merely wished to enable various AI applications. Now it was becoming clear that this was a monumental task that required revision of many other laws and first large-scale compliance and audit mechanisms could be observed on the horizon.

In the second year of its mandate the group went ahead and developed its third deliverable in 2020, a practical tool called Assessment List for Trustworthy AI (ALTAI).⁵⁷ This document builds on the earlier second deliverable by expanding the definitions of the 7 key requirements for Trustworthy AI. The document is structured as an audit questionnaire. For each of the key requirements it sets out series of generalised questions targeted at the AI producer/manufacturer.

For example:

Human Agency and Oversight:

- *Did you ensure a ‘stop button’ or procedure to safely abort an operation when needed?*
- *Did you take any specific oversight and control measures to reflect the self-learning or autonomous nature of the AI system?*

Technical Robustness and Safety:

- *Did you define risks, risk metrics and risk levels of the AI system in each specific use case?*

Privacy and Data Governance

- *Is your AI system being trained, or was it developed, by using or processing personal data (including special categories of personal data)?*

Transparency

- *Did you put in place measures that address the traceability of the AI system during its entire lifecycle?*

Diversity, Non-discrimination:

- *Does the AI system potentially negatively discriminate against people on the basis of any of the following grounds? (non-exhaustively): sex, race, colour, ethnic or social origin, etc.*

⁵⁷ AI HLEG – Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. Digital-strategy.ec.europa.eu. [online]. [cit. 2020-07-17]. Available from: <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>

Societal and Environmental Well-being

- *Where possible, did you establish mechanisms to evaluate the environmental impact of the AI system's development, deployment and/or use (for example, the amount of energy used and carbon emissions)?*

Accountability

- *Did you consider establishing an AI ethics review board or a similar mechanism to discuss the overall accountability and ethics practices, including potential unclear grey areas?*

There are nearly over 100 questions that would in total generate about 100 pages of compliance-based text for auditors to go through. However, it does not establish benchmarks or correct answers. It neither proposes any specific law enforcement mechanisms how to cross reference expected answers with evidence, or even what type or quality of evidence would be admissible. It is here that the fears of overregulation are grounded.

The final deliverable, and perhaps the final nail in the coffin for the AI HLEG were the Sectoral Considerations on Policy and Investment Recommendations for Trustworthy AI⁵⁸. In this report the group chose three areas (Public Sector, Healthcare, Manufacturing and Internet of Things) and within context of each area provided its thoughts on the type of problems that might arise. Unlike the previous deliverables, the final one was very brief and lacked any new meaningful input. Perhaps that is why the European Commission decided to disband the group in July 2020⁵⁹ and transform it into a new consultation body called AI Alliance.

These four deliverables were however crucial in laying the foundations for the topics and methods that eventually found their way into the European Commission's White Paper on Artificial Intelligence⁶⁰. Alongside that the Commission updated its original Coordinated Plan on Artificial Intelligence from 2018 into a new version which they released in 2021⁶¹. The new Coordinated Plan of 2021 invited the member states to publish their own national strategies. These

⁵⁸ AI HLEG – Sectoral Considerations on Policy and Investment Recommendations for Trustworthy AI. Futurium.ec.europa.eu. [online]. [cit. 2020-07-23]. Available from: <https://futurium.ec.europa.eu/en/european-ai-alliance/document/ai-hleg-sectoral-considerations-policy-and-investment-recommendations-trustworthy-ai>

⁵⁹ AI HLEG. Digital-strategy.ec.europa.eu [online]. [cit. 2021-09-17]. Available from: <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai>

⁶⁰ EC - White Paper on Artificial Intelligence: a European approach to excellence and trust. Ec.europa.eu. [online]. [cit. 2020-02-19]. Available from: https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

⁶¹ EC – Coordinated Plan on Artificial Intelligence 2021 Review. Digital-strategy.europa.eu. [online]. [cit. 2021-04-21]. Available from: <https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review>

strategies were used as a foundation for research published by European Commission's Joint Research Centre (JRC) and the OECD's Science Technology and Innovation Directorate⁶².

Meanwhile, AI Alliance – an online forum with over 4000 members representing academia, business and industry, civil society, EU citizens and policymakers – met for the first time in June 2019 on the 1st European AI Alliance Assembly and then again in October 2020 and September 2021⁶³. Their function is to provide relevant feedback for policy documents.

STOA is a panel for the Future of Science and Technology (STOA), and managed by the Scientific Foresight Unit, within the Directorate-General for Parliamentary Research Services (EPRS) of the Secretariat of the European Parliament. Being separate from the European Commission, this body created their own study called *The Ethics of Artificial Intelligence: Issues and Initiatives*.⁶⁴ As is evidenced from the amended revision by the European Parliament⁶⁵, while there is a significant certain overlap and agreement on the core issues, the position of the European Parliament is much stricter towards AI and its manufacturers. For example, it negatively positions itself towards the utilisation of broad biometric identification and surveillance, and wants to include social media algorithms.

3.5. Values extracted from AI proto-regulation and their mirror image in legislation

A value may be defined as a parameter that expresses the degree of importance of an object in comparison to other objects. It would be difficult to claim that the system of values is only strictly hierarchical and never changes with time and context. Each moral agent has its own set of values, which create a *a priori* moral framework through which everything is perceived and all actions are processed and judged.

Law in context of this chapter may be viewed as a superimposed enforceable system of norms and principles applicable to moral agents, which are based on a system of pre-agreed values that are held by the legislator - i.e. the ultimate supreme authority.

⁶² EC/OECD – New report looks at AI national strategies progress and future steps publications.jrc.ec.europa.eu [online]. [cit. 2021-09-16]. Available from: <https://publications.jrc.ec.europa.eu/repository/handle/JRC122684>

⁶³ EC – The First European AI Alliance Assembly. Digital-strategy.europa.eu [online]. [cit. 2019-06-26]. Available from: <https://digital-strategy.ec.europa.eu/en/events/first-european-ai-alliance-assembly>

⁶⁴ STOA – The ethics of artificial intelligence: Issues and initiatives. Brussels: Panel for the Future of Science and Technology (STOA), Scientific Foresight Unit, Directorate-General for Parliamentary Research Services (EPRS) of the Secretariat of the European Parliament. ISBN: 978-92-846-5799-5 Available from: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU\(2020\)634452_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf)

⁶⁵ EP – Draft of Compromise Amendments on the Draft Report Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts. Europarl.europa.eu. [online]. [cit. 2023-05-16]. Available from: <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>

It is also worthwhile to make a distinction between values and features. Values are the most generalised ideals of what is deemed important, whereas features are more specific, function oriented and closely resemble a proto-right or proto-obligation.

It is curious that given that this is regulation about AI applications, one could reasonably expect it to be a document more of a technical nature, whereas in reality most of these values are revolving around human rights. This translates into the legislation in multiple ways.

First, some AI practices are prohibited because they are deemed to resemble a direct attack on critical human rights.⁶⁶ Such as placing on the market an AI system that is manipulative or exploitative in regards to disadvantaged individuals or introduces strict limits to utilisation of real-time biometric or facial recognition AI applications.

Secondly, it puts human security and well-being at the centre of the officially institutionalised and privately delegated risk analysis. The institutional risk analysis takes place when Commission considers a decision whether or not to add the AI system onto a watchlist of Annex III⁶⁷. The privately delegated risk analysis happens throughout entire lifecycle of the AI system and is a continuous iterative process⁶⁸.

Third, a series of brand new rights assigned directly to users of the AI application. These will be covered in detail in later chapter. They are not written in the act explicitly as you would find in GDPR⁶⁹, but hidden latently in the obligations of the providers and manufacturers.

The White paper on AI encourages this narrative by explicitly pointing out that the Union should strive forward to human centric legislation, for the reasons so as to create a technology that is trustworthy, safe and respects fundamental rights.⁷⁰

All of these values point quite clearly to set of features that are expected of an EU-compliant AI system. These demands of features are translated in the legislation in a series of obligations that affect anyone even remotely involved with the AI application⁷¹, that includes the providers, manufacturers, importers, distributors and users. Here are some examples of demanded features:

- Continuous quality management system monitoring the operation of the AI system

⁶⁶ Article 5, EC – Proposal for a Regulation of the European Parliament and of the Council on Artificial Intelligence

⁶⁷ Article 7, EC – Proposal for a Regulation of the European Parliament and of the Council on Artificial Intelligence

⁶⁸ Article 9, EC – Proposal for a Regulation of the European Parliament and of the Council on Artificial Intelligence

⁶⁹ EP – Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). eur-lex.europa.eu [online]. [cit. 2023-02-05]. Available from:

<https://eur-lex.europa.eu/legal-content/EN-CS/TXT/?from=CS&uri=CELEX%3A32016R0679>

⁷⁰ EC – White Paper on Artificial Intelligence: a European approach to excellence and trust.

⁷¹ Article 13, 14, 16-29, EC – Proposal for a Regulation of the European Parliament and of the Council on Artificial Intelligence

- Technical documentation detailing the inner functioning of the AI system
- Automatically generated logs capturing the function of the AI system
- Procedure of corrective measures and effective human oversight
- Duty of provision of information and explanation

As the usage of AI systems become more commonplace, we may expect the list of obligations grow ever larger and be applied to an ever-increasing list of products and services. It should be noted that information presented as valid as of time of writing this paper. More accurate information should be gathered for the version of the legislation that will be in effect.

3.6. Calls for decisive action and the European legislative process

The legislative roadmap for AI has been severely delayed due to COVID-19 pandemic and War on Ukraine. The ambitions of the Coordinated Plan on AI were slightly relaxed in light of these developments. The European Commission unveiled the proposal for the new Artificial Intelligence Act in April 2021⁷². This is the standard legislative procedure for the EU legislation, where the only entity in the EU endowed with legislative initiative is the European Commission.⁷³ Year later, the Council has adopted its common position on the AI Act in December 2022.⁷⁴ Few months later in May 2023 the European Parliament took the initiative and released their version of compromise amendments to the Proposal for the Regulation on AI⁷⁵. The discussions were lead by two Committees, Internal Market and Consumer Protection (IMCO) and the Committee on Civil Liberties, Justice and Home Affairs (LIBE) under a joint committee procedure. It is expected that the European Parliament will vote the IMCO/LIBE report in June 2023 during their plenary session, which will be the final step needed before the interinstitutional negotiations that will eventually produce the final joint text of the regulation.

These negotiations however need to be quick because we are expecting elections for the European Parliament in May 2024⁷⁶. Should we assume that all three institutions will approve of

⁷² EC – Proposal for a Regulation of the European Parliament and of the Council on **Artificial Intelligence**. Eur-lex.europa.eu. [online]. [cit. 2021-04-21]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>

⁷³ TOMÁŠEK, Michal.; TÝČ, Vladimír a kol.. Právo Evropské unie. 2. vyd.. Praha : Leges. 2017. 496 s.. 78-80-7502-184-7. str. 196

⁷⁴ EP – Proposal for a Regulation on a European approach for Artificial Intelligence In “A Europe Fit for the Digital Age”, europarl.europa.eu. [online] [cit. 2023-06-03] Available from: <https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence>

⁷⁵ EP – Draft of Compromise Amendments on the Draft Report Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts. Europarl.europa.eu. [online]. [cit. 2023-05-16]. Available from: <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>

⁷⁶ EP – European elections 2024, multimedia.europa.eu [online] [cit. 2023-05-23]. Available from: https://multimedia.europarl.europa.eu/en/package/european-elections-2024_23001

the final text before the elections, due to the scope of responsibilities and industries affected, we can expect the standard 2 year implementation period⁷⁷ before the new rules take effect. Not to mention that we may expect certain delays in the implementation period deadline as we have seen with GDPR's implementation, where the ultimate deadline was postponed from 2014 to 2015⁷⁸.

Therefore the realistic expectation of the new rules being in effect is 2026-2027. Given the rapid rise in AI applications we have seen in the last few months it seems unfortunate that there will be a period of time of almost 5 years, where the AI applications will be limited only by existing laws.

3.7. The structure of the regulation and comparison of positions for legislative trilogue

The original proposal by the European Commission⁷⁹ sets the stage by laying down a regulation that closely resembles the NIS2⁸⁰ directive due to its internal structure.

First the regulation establishes brand new the definitions⁸¹. Artificial intelligence system is defined as 'software that is developed by techniques listed in Annex I'. It must be stressed that the Annex I can be updated by the Commission at any time to enlarge the scope of regulated entities.⁸² These techniques mentioned in Annex I⁸³ presently are:

- *Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;*
- *Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;*
- *Statistical approaches, Bayesian estimation, search and optimization methods.*

What is quite surprising is that AI has traditionally been defined as being one of the first two points. The inclusion of all statistical approaches is understandable because of possible fears

⁷⁷ EC – Proposal for a Regulation of the European Parliament and of the Council on Artificial Intelligence. Eur-lex.europa.eu. [online]. [cit. 2021-04-21]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>, Article 85

⁷⁸ PrivacyTrust – Delays to the New EU Data Protection Law Implementation privacytrust.com [online] [cit. 2023-06-04] Available from: <https://privacytrust.com/delays-to-the-new-eu-data-protection-law-implementation-2/>

⁷⁹ EC – Proposal for a Regulation of the European Parliament and of the Council on Artificial Intelligence. Eur-lex.europa.eu. [online]. [cit. 2021-04-21]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>

⁸⁰ (L10) EP – Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive) (eur-lex.europa.eu [online]. [cit. 2023-03-24]. Available from: <https://eur-lex.europa.eu/eli/dir/2022/2555>

⁸¹ Idem. Article 3

⁸² Idem. Article 4

⁸³ EC – Annex to the Proposal for a Regulation of the European Parliament and of the Council on Artificial Intelligence. Eur-lex.europa.eu. [online]. [cit. 2021-04-21]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>

that providers of AI applications might downplay their nature to being mere statistical models. The downside of the inclusion however is that it expands the scope of the regulation to nearly every system that processes any data, because nearly every software either searches or analyses data.

This inclusion of statistical approaches was not welcomed by the common position of the Council, that proposes eliminating this category and keeping only the first two categories⁸⁴.

However the fact alone that something is defined as AI does not mean it will fall under the strict regulatory demands. The regulation, though not explicitly, but rather implicitly then establishes 4 levels of AI risks.

- Unacceptable risk AI – These types are banned outright
- High-risk AI – These types are heavily scrutinised by a vast compliance procedure
- Limited risk AI – These types have to adhere to only a limited set of very basic obligations
- Minimal risk AI – These types are not regulated at all and are free of additional legal obligations, though other existing legislation has to be taken into account.

Unacceptable risk AI is defined not by its technical nature, but rather by its use, specifically from the perspective of basic human rights and liberties. For example,⁸⁵:

- *AI system that deploys subliminal techniques beyond person's consciousness in order to materially distort a person's behaviour*
- *AI system that exploits any of the vulnerabilities of a specific group of persons*
- *AI systems by public authorities or on their behalf for the evaluation or classification of the trustworthiness of natural persons, with the social score leading to detrimental or unfavourable treatment*
- *AI system that uses 'real-time' remote biometric identification systems in publicly accessible spaces*

There is more nuance to the banned AI practices and will surely be subject to a heated debate in the trilogue.

High-risk AI is defined by two methods. First it is a product, covered by the Union harmonisation legislation listed in Annex II⁸⁶. These are 12 pieces of aforementioned legislation.

⁸⁴ EP – Proposal for a Regulation on a European approach for Artificial Intelligence In “A Europe Fit for the Digital Age”, [europarl.europa.eu](https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence). [online] [cit. 2023-06-03] Available from: <https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence>

⁸⁵ EC – Proposal for a Regulation of the European Parliament and of the Council on Artificial Intelligence. [eur-lex.europa.eu](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206). [online]. [cit. 2021-04-21]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>, Article 5

⁸⁶ Idem. Annex II

These are products such as medical machinery⁸⁷, in vitro diagnostic medical devices⁸⁸, machinery⁸⁹, lifts⁹⁰, protective systems intended for use in potentially explosive atmospheres⁹¹, toys⁹², recreational boats⁹³, planes⁹⁴, agricultural and forestry vehicles⁹⁵, trains⁹⁶, vehicles⁹⁷

Second it falls into a category of certain use-cases covered by list in Annex III⁹⁸, these are:

- *Biometric identification and categorisation of persons.*
(Will be sought to be banned by the European Parliament outright⁹⁹)
- *Management of critical infrastructure such as road transport and delivery of public services.*
- *Education and training, particularly for access, allocation and assessment decisions.*
- *Employment, including recruitment, promotion, termination, assignment and performance appraisal.*
- *Public and private essential services, such as benefit eligibility evaluation, credit evaluation, and emergency dispatch.*

⁸⁷ Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (OJ L 117, 5.5.2017, p. 1);

⁸⁸ Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU (OJ L 117, 5.5.2017, p. 176).

⁸⁹ Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC (OJ L 157, 9.6.2006, p. 24) [as repealed by the Machinery Regulation];

⁹⁰ Directive 2014/33/EU of the European Parliament and of the Council of 26 February 2014 on the harmonisation of the laws of the Member States relating to lifts and safety components for lifts (OJ L 96, 29.3.2014, p. 251);

⁹¹ Directive 2014/34/EU of the European Parliament and of the Council of 26 February 2014 on the harmonisation of the laws of the Member States relating to equipment and protective systems intended for use in potentially explosive atmospheres (OJ L 96, 29.3.2014, p. 309);

⁹² Directive 2009/48/EC of the European Parliament and of the Council of 18 June 2009 on the safety of toys (OJ L 170, 30.6.2009, p. 1);

⁹³ Directive 2013/53/EU of the European Parliament and of the Council of 20 November 2013 on recreational craft and personal watercraft and repealing Directive 94/25/EC (OJ L 354, 28.12.2013, p. 90);

⁹⁴ Regulation (EC) No 300/2008 of the European Parliament and of the Council of 11 March 2008 on common rules in the field of civil aviation security and repealing Regulation (EC) No 2320/2002 (OJ L 97, 9.4.2008, p. 72).

⁹⁵ Regulation (EU) No 167/2013 of the European Parliament and of the Council of 5 February 2013 on the approval and market surveillance of agricultural and forestry vehicles (OJ L 60, 2.3.2013, p. 1);

⁹⁶ Directive (EU) 2016/797 of the European Parliament and of the Council of 11 May 2016 on the interoperability of the rail system within the European Union (OJ L 138, 26.5.2016, p. 44).

⁹⁷ Regulation (EU) 2018/858 of the European Parliament and of the Council of 30 May 2018 on the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles,

⁹⁸ EC – Proposal for a Regulation of the European Parliament and of the Council on Artificial Intelligence. Eur-lex.europa.eu. [online]. [cit. 2021-04-21]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>, Annex III

⁹⁹ EP – Draft of Compromise Amendments on the Draft Report Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts. Europarl.europa.eu. [online]. [cit. 2023-05-16]. Available from: <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>

- *Law enforcement, including individual risk assessment, detection of emotional states, detection of false evidence, assessment of the reliability of evidence, crime prediction, profiling, and identification of complex data patterns.*
- *Migration, asylum and border control, such as emotional state detection, risk assessment, document authentication and application screening.*
(will be sought to be banned by the European Parliament outright¹⁰⁰)
- *Administration of justice and democratic processes, specifically assisting the judiciary in researching, interpreting and applying the law.*

Additionally, the European Parliament would like to add:

- *AI systems intended to be used for influencing the outcome of an election or referendum or the voting behaviour*
- *AI systems intended to be used by social media platforms that have been designated as very large online platforms*

While the Commission's initial draft automatically categorizes as high-risk all systems falling in areas mentioned in Annex II or use cases mentioned in Annex III, the European Parliament¹⁰¹ would like to add additional requirement that the systems must pose a 'significant risk' to qualify as high-risk, that way significantly reducing the scope of the regulation. Though what would qualify as 'significant risk' is opening doors to interpretation, which leads to uncertainty. The majority of the regulatory text and its compliance requirements are directed at high-risk AI systems, therefore falling within scope of this category will be dreaded fate of all AI providers.

Third category that is subject to a very reduced set of regulatory ambitions is Limited Risk AI. This group is composed of those products or entities that fall within the definition of AI, but not qualify to fall within definition of High-Risk AI. The responsibilities of providers of Limited Risk AI¹⁰² are as follows:

- *AI systems intended to interact with natural persons are designed such a way that natural persons are informed that they are interacting with an AI system.*

¹⁰⁰ EP – Draft of Compromise Amendments on the Draft Report Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts. [Europarl.europa.eu](https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence). [online]. [cit. 2023-05-16]. Available from: <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>

¹⁰¹ Idem.

¹⁰² EC – Proposal for a Regulation of the European Parliament and of the Council on Artificial Intelligence. [eur-lex.europa.eu](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206). [online]. [cit. 2021-04-21]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>, Article 52

- *Users of an emotion recognition system or a biometric categorisation system shall inform of the operation of the system the natural persons exposed thereto.*
- *Users of an AI system that generates or manipulates image, audio or video content ('deep fake'), shall disclose that the content has been artificially generated or manipulated.*

That covered the definition segment of the regulation, though the aforementioned explanation is not exhaustive, it is sufficient for understanding the general outlook. Besides, the details of definitions are still subject to change during the legislative process.

The next part of the regulation goes into the specific responsibilities targeted at the High-Risk AI systems¹⁰³. Articles 8 to 15¹⁰⁴ deal with specific built-in features that the legislator demands that High-Risk AI systems include. These features have been extrapolated from the values that preceded in the aforementioned protoregulatory documents. For example:

- *Risk management system*
- *Data and data governance*
- *Technical documentation*
- *Record-keeping*
- *Transparency*
- *Human oversight*
- *Accuracy, robustness and cybersecurity*

Following that is a chapter dedicated to the responsibilities of providers, importers, distributors and users of High-Risk AI systems. In this chapter, besides going into specifics of the mentioned expected built-in features, the details of compliance procedure and conformity assessment are explained. Providers are expected to resolve majority of those demands, while importers and distributors merely have the secondary duty to check whether the provider has sufficiently fulfilled his or her duties. While users are expected to follow their training, proper procedure, enter only valid input data, report any error encountered during operation and stop the AI system should it exceed limits of its expected behaviour.

Though there is an interesting measure in Article 28¹⁰⁵, where any distributor, importer, user or other third-party shall be considered a provider for the purposes of this Regulation and thus

¹⁰³ EC – Proposal for a Regulation of the European Parliament and of the Council on Artificial Intelligence. Eur-lex.europa.eu. [online]. [cit. 2021-04-21]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>, Article 8 et. seq.

¹⁰⁴ Idem. Article 8 et. seq.

¹⁰⁵ Idem. Article 28

shall be subject to the obligations of the provider under Article 16¹⁰⁶, in any of the following circumstances:

- they place on the market or put into service a high-risk AI system under their name or trademark;
- they modify the intended purpose of a high-risk AI system already placed on the market or put into service;
- they make a substantial modification to the high-risk AI system.

What qualifies as substantial modification though is – once again - left unanswered, which is another point of possible critique, because it raises uncertainty for the users of High-Risk AI systems. The difference between the responsibilities of provider and user is substantial and costs associated with compliance dramatically astronomical, therefore we may expect this to be another point of contention.

Next the regulation lays down the framework for the upcoming institutions across the EU that will be responsible for setting up and carrying out the procedures for the assessment of conformity. Each member state is expected to designate or establish a notifying authority¹⁰⁷ that will report to the new central EU institution, the European Artificial Intelligence Board¹⁰⁸, or ‘AI Office’ as proposed by the European Parliament. The national notifying authorities will be issuing Certificates for EU compliant High-Risk AI systems, such systems will need to carry the CE marking of conformity. The national notifying authorities will communicate with the national competent authorities¹⁰⁹, whose role is to ensure the application and implementation of this Regulation. Conformity assessment bodies will qualify through this Regulation to become notified bodies that will then perform third-party conformity assessment activities, including testing, certification and inspection¹¹⁰.

A non-negligible portion is dedicated to a specific regime of AI regulatory sandboxes, though I remain doubtful over its meaningfulness in practice.

Finally, the law enforcement expects that the public authorities will have direct access to data, documentation, training data and application programming interfaces (‘API’) of the High-Risk AI systems, though naturally bound by administrative confidentiality.

¹⁰⁶ Idem Article 16

¹⁰⁷ EC – Proposal for a Regulation of the European Parliament and of the Council on Artificial Intelligence. Eur-lex.europa.eu. [online]. [cit. 2021-04-21]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>, Article 30 et. seq.

¹⁰⁸ Idem. Article 56

¹⁰⁹ Idem Article 59

¹¹⁰ Idem. Article 33

In regards to penalties, they are quite substantial. Non-compliance with Article 5 (prohibited AI practices) may result in a fine of up to 30 000 000 EUR or, if the offender is company, up to 6 % of its total worldwide annual turnover for the preceding financial year, whichever is higher. Should the non-compliance take other form besides the breach of Article 5, a fine of up to 20 000 000 EUR or, if the offender is a company, up to 4 % of its total worldwide annual turnover for the preceding financial year, whichever is higher can be expected.

3.8. Methodology of regulation (Law)

In this chapter we will take a look at the generalised methods of drafting regulation available to the legislator in regards to the topic of AI. This chapter has been heavily influenced and inspired by thoughts and work of Nicolas Petit and Jerome De Cooman from the Robert Schuman Centre for Advanced Studies at the European University Institute.

Nicolas Petit's ideas are relevant to our topic for many reasons. He was one of the first academics to start to write consistently, conceptually and thoughtfully about the topic of Law and AI, for example his paper "Law and Regulation of Artificial Intelligence and Robots: Conceptual Framework and Normative Implications" published in March 2019, where he correctly assessed that segregated disciplinary bottom-up approach is by definition full of contradictions in the grand scheme of things. He described the dynamics of misguided attempts to hastily fix issues relating to accountability in the 2017 European Parliament resolution on Civil law rules on robotics¹¹¹ that was devoid of any methodological thought or more grandeur design behind it. Furthermore, he introduces a very effective method of thinking about regulatory methods, which is founded in the assessment of the need for the regulation through the lenses of market dynamics and typology of externalities, be it positive or negative ones.

Additionally on top of his academic work, he was part of the AI HLEG¹¹² group and worked on many of its deliveries, such as Ethics Guidelines (2019), Policy Recommendations (2019), Assessment List for Trustworthy AI (2020) and Sectoral Policy Recommendations (2020). Many of the ideas captured within these documents have found their way into official documents that followed, such as the White Paper on Artificial Intelligence – A European approach to excellence and trust. Subsequently in 2021 the European Commission has published several proposals such as Proposal for "Regulation on Machinery Products", or the "Regulation on Artificial Intelligence" and finally Proposal for "Declaration of Digital Principles". These will be explored in detail in later chapter.

¹¹¹ European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)).

¹¹² AI HLEG – Independent High-Level Expert Group on Artificial Intelligence (set up by the European Commission)

Therefore, understanding the ideas in the paper of Models of Law and Regulation for AI are crucial stepping stone in order to understand the subsequent attempts by the European Commission because they follow along the lines of the framework that was first introduced there.

3.9. Introduction to Regulation Models for AI

The paper goes into 5 models of regulation on AI in total, where the last fifth one is a combination of all the previous 4 models which are applied only in specific situations of certain predefined market externalities. Below is a brief and simplified explanation of these models for the purposes of subsequent analysis in this paper. For more detailed understanding of these models and their fine details I recommend reading the original paper.

	Black Letter Law	Emergent	Ethical	Risk Regulation
Timing	Reactive	Proactive	Proactive	Proactive
Discussion	Descriptive	Normative	Normative	Normative
Approach	Statutory and doctrinal interpretation of de lege lata	Normative creative of de lege ferenda	Teleological when deontological; Ontological when consequentialism	Cost-benefit analysis with possible precautionary principle
Issues	Irrelevance	Redundance	Ethics lobbying Ethical relativism Ethics shopping	Knee-jerk regulation ¹¹³

The European Union has taken in its Proposal for a Regulation on Artificial Intelligence very clearly the Risk Regulation approach, similarly to NIS2 Cybersecurity Directive¹¹⁴. Most visibly it is manifested by the scope of regulation because, as defined by currently proposed Article 6(2)¹¹⁵, it specifically addresses High-Risk AI systems only, which are specified further in an Annex III.

These High-risk AI systems are identified not by their scientific topology or data model archetype, but rather by their function and use-case and risk associated with that use-case. That is an exceedingly broad scope.

What is quite interesting is the fact that whereas the NIS2 Directive targets specific high-profile industries that are deemed essential or important, the AI regulation targets specific use cases. Therefore, the scope of AI regulation is broader. Should the current proposal be

¹¹³ A subtle modification and simplification of Table 1 from Chapter 1 of (A2) Nicolas Petit, Jerome De Cooman – Models of Law and Regulation for AI. Robert Schuman Centre for Advanced Studies Research Paper No. RSCAS 2020/63, EUI Department of Law Research Paper papers.ssrn.com. [online]. [cit. 2020-10-08]. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3706771

¹¹⁴ EP – 2022/2555 (NIS 2 Directive)

¹¹⁵ EC – Proposal for a Regulation of the European Parliament and of the Council on Artificial Intelligence

implemented as is, all entities subject to NIS2 regulation will have to implement their own AI compliance programs for their legal continuous operation, perhaps as an extension of their NIS2 compliance.

Another sign that the proposed AI regulation is taking the risk regulation approach is the types of responsibilities and law enforcement mechanisms it applies.

The advantage of risk-based approach is its flexibility, this flexibility is further reinforced in the empowerment of legislation to update the scope of the regulation through annex updates.

The disadvantage of risk-based approach is uncertainty and lack of clarity in regards to where the line is drawn between what is out of scope and what is not of the regulation. Furthermore, additional disadvantage is that should someone somewhere incorrectly gauge the risk-factor, harm may not be pre-emptively prevented. Although due to nature of AI applications, persistent updates and live service aspect, there is no other way to approach this from functional regulatory perspective.

In conclusion I believe the risk-based approach was the only possible meaningful method to deliver a piece of legislation what will scale and keep up with the rapidly evolving AI applications and their deployment over many fields.

It should be noted that there are other accompanying pieces of laws that deal with other aspects of AI related issues, such as the question of liability, of which there are two separate directives, the AI Liability Directive¹¹⁶ and the revamped Product Liability Directive¹¹⁷, which are taking a different approach more akin to traditional of product-safety style regulation, or establishment of strict framework of rights and responsibilities alike to traditional civil responsibility. Additionally, we ought to mention there is already in effect the European Declaration on Digital Rights and Principles for the Digital Decade¹¹⁸, which functions as a declaration of regulatory zeal and intention for the digital landscape, as well as high-level guidance document that should be used for interpretation of other legislation, both old and new, to be in line with.

We have therefore assessed that the proposed European Regulation on AI has taken the Risk Regulation approach, by focusing its attention on the function and use-case of the AI application rather than the nature of the AI architecture in relation to aforementioned AI topology or data model archetypes.

¹¹⁶ EP – Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive)

¹¹⁷ EP – Proposal for a Directive of the European Parliament and of the Council on liability for defective products

¹¹⁸ EP – European Declaration on Digital Rights and Principles for the Digital Decade

4. What is (should be) Artificial intelligence from regulatory perspective?

4.1. Lessons learned from the human condition

There are so many types of AI it is difficult to approach the topic in a meaningful way without succumbing to generalisations which become so abstract and high-level that they stop being meaningful. We have covered different approaches to definitions of man, mind and personhood. All approaches have their place and rationale. Some of them are more useful than others. What would happen to the quality of said definitions if we simply swapped human or man for an AI?

Ethics

According to Bible, God made man in his likeness, therefore the creations of man by implication also.

According to Buddhism, AI may be a mindstream, state of memories in a continuum.

Kant has already generalised his definition by addressing his categorical imperatives to the sentient beings. His criterions are self-reflective capacities and rationality. However, we conceded there was a spectrum and couple of exceptions.

Science

According to biology, AI can be considered as an entity that has „complex brain“ and whose activity leads to creation of advanced tools and language.

According to psychology, AI may have consciousness if defined as awareness of internal and external existence, for example via sensors, just like we are aware with our biological sensors.

Economics highlights the capacity to act rationally and most optimally. It is self evident that AIs are created for that purpose.

Law

Most legal systems clearly qualify AI as a tool. But it would appear that such logical and simple conclusion alone may not be functional in all cases, and since law is all about being functional, it would be a bad law. We are met with many problems and we intuitively feel there is something to do here but we do not know where to quite start.

4.2. AI according to AI scientists

We may all reasonably assume that the best for the job to define what is an AI are the very people who created it, AI scientists and experts in computer science. When diving into relevant literature, of which there is a lot, one might encounter following terms: Strong AI, Weak AI, Machine learning (supervised and unsupervised learning based on data), Reinforcement learning

(random trial and error and pursuit of a reward function), Deep learning (aspect of multiple processing layers). It would be beyond the scope of this paper to explain what these terms mean and I refer to the body of literature present.

The bottom line is that these are terms that describe the inner functioning and the method of assembly and design of architecture, rather than functional definition that is useful for legislator.

The takeaway for regulator from this approach is that there are two types of AI methods which differ in auditability. First, are the systems based on “if this then that” statements (also known as expert systems), where decision-making process can be clearly traced from output to input, given enough time and effort. Second, are the systems which have no explicit “rules or statements” but instead are goal oriented or reward function oriented. The programmer of the AI lets the AI itself to come up with the best method of how to arrive to a desired result, without telling it how to do it, and in some cases the programmer will not even tell the AI what the desired result is and leaves it guessing. This approach can be best exemplified in neural network-based AI systems.

Another take also present in the literature on computer science is a more conceptual approach, exemplified very well by Stuart Russel in his book *Human compatible: artificial intelligence and the problem of control*¹¹⁹. Russel’s take is more useful because he approaches the topic of AI from a control perspective. In his book he lays out 3 stages of AI:

- artificial narrow intelligence (ANI),
- artificial general intelligence (AGI),
- artificial super intelligence (ASI).

The long and the short of the “AI control problem”, also known as “Paperclip apocalypse problem”¹²⁰ or as “Instrumental convergence” is that there is a very valid danger that unshackled decision-making agent tasked with a relatively benign and harmless goal may, if not stopped or controlled, inevitably cause rampant damage of unforeseen scale. Such as machine tasked with making as many paperclips as possible, which results in AI taking over the world to capture all resources of Earth and switch all factory production to making paperclips. If there was a “stop” big red button, the AI would do everything in its power to prevent anything or anyone pressing it, because it would go against its primary goal – making as many paperclips as possible. It is a behavioural paradox and the very essence of the control problem.

¹¹⁹ RUSSEL, S (October 8, 2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. United States: Viking. ISBN 978-0-525-55861-3.

¹²⁰ GANS, J S (2017), “Self-Regulating Artificial General Intelligence”, arXiv:1711.04309.

4.3. AI according to Policy makers

Not to overstay welcome in this chapter, which could be very much endless, I will now introduce a document that I believe was the first body of text that kicks off the discussion for the regulator very well. It is a publication by Joint Research Centre (JRC), the European Commission's science and knowledge service, titled AI Watch – Defining Artificial Intelligence 2.0.¹²¹ The document introduces a taxonomy for AI and operational definition of AI. It is an amalgamation of research that narrows down both scientific attempts at definition (as introduced above) and protoregulatory attempts at definition (for example the achievements by HLEG, more on that later).

The publication identifies these main features of AI:

- **Perception** of the environment, including the consideration of the real-world complexity
- **Information processing**: collecting and interpreting inputs (in form of data)
- **Decision making** (including reasoning and learning): taking actions, performance of tasks including adaptation, reaction to changes in the environment) with certain level of autonomy
- **Achievement of specific goals**: this is considered as the ultimate reason of AI systems

It goes without saying that the reason why AI is regulated is for its decision-making ability and by extension the ability to affect the world.

¹²¹ EC – AI Watch. Defining Artificial Intelligence 2.0. publications.jrc.europa.eu [online]. [cit. 2021-10-29]. Available from: <https://publications.jrc.ec.europa.eu/repository/handle/JRC126426>

The publication in its 125 pages lays out a taxonomy of various AI types and assigns definitions and explanations. To summarise their findings here is a table from the publication that gives outlines of topics of AI subdomain disciplines:

Table 1. AI domains and subdomains constituting one part of the operational definition of AI

		AI taxonomy	
		AI domain	AI subdomain
Core	Reasoning		Knowledge representation
			Automated reasoning
			Common sense reasoning
	Planning		Planning and Scheduling
			Searching
			Optimisation
	Learning		Machine learning
	Communication		Natural language processing
	Perception		Computer vision
			Audio processing
Transversal	Integration and Interaction		Multi-agent systems
			Robotics and Automation
			Connected and Automated vehicles
	Services		AI Services
	Ethics and Philosophy		AI Ethics
			Philosophy of AI

Table 1. AI domains and subdomains within European AI Taxonomy¹²²

The publication makes it very clear that these suggested subdomains are all related and influence one another. This chart is a work in progress and will be regularly updated as progress in the field of AI advances.

4.1. Analysis of the functional features of AI and legislative recommendations

From regulatory perspective it would be useful to give more robust description of these features because these features are not binary, but rather, they are on a scale. Moreover, in some scenarios of approaches to legislation (specifically risk-based approaches) the degree to which certain feature of AI is realised has direct impact on whether that specific AI application will be within scope of the regulation or not. In following subchapters these degrees are further explored with examples.

¹²² EC – AI Watch. Defining Artificial Intelligence 2.0. publications.jrc.europa.eu [online]. [cit. 2021-10-29]. Available from: <https://publications.jrc.ec.europa.eu/repository/handle/JRC126426>

4.1.1. Epistemology (Perception and acquisition of data and knowledge)

The term epistemology is chosen here purposefully because it signifies a more meta-approach to the theory of knowledge and acquisition of thereof. In practical sense the feature of AI to capture information about the world around them is procured by sensors, which it is endowed with. It would be incomprehensive to reduce the potential palate of senses to the mere basic human senses (sight, hearing, taste, touch, smell). There are and will be created many more types of sensors that go beyond these, some inspired in nature (for example echolocation of bats or sense of magnetic field by bees), some totally new and artificial (detection of gravitational waves or other exotic particles). Each of these sensors has a quality to it. An eye of a snake, human and an eagle is quite different, but it can go beyond that. We can consider sensors that can capture more wavelengths than visible light, such as infrared cameras. We can also consider stacking multiple of those sensors to different places to create a more omniscient presence such as is case with CCTV cameras.

For regulator when writing legislation, or judge when deciding a case, or any human in any position, it is relevant whether AI has the capacity to observe certain fact, and to what degree of quality and certainty, and if it did not said capacity, it is relevant to ask why it was not equipped with said ability, or why has said capacity failed. The obvious ramifications to these questions are for example questions of responsibility, questions of reliability of alleged fact, and so on. Therefore, it would be advisable if for example a legislator introduced a certain quality criterion for sensory abilities as a precondition for using AI applications within certain high-risk environments or at least a method to ascertain the degree of reliability and quality of sensors.

All of these sensors together generate data, structured in a database. Data and databases themselves are already subject to regulation and it is an ongoing process. The go-to example of such regulation is of course GDPR¹²³, but long before that we had and still have in force the Directive on the legal protection of databases¹²⁴. Quite recently there has been a new legislation that facilitates data-sharing held by public bodies, so called Data Governance Act.¹²⁵ Currently as a heavily contested proposal for regulation there is also a Data Act, which establishes right to users of products in certain cases to demand data held by private companies.¹²⁶

¹²³ EP – Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

¹²⁴ EP – Directive (EU) 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases

¹²⁵ EP – Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation

¹²⁶ EP – Proposal for a Regulation of the of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)

All of these are relevant regulations that while they do not have necessarily AI in title, nonetheless they directly apply to it. Some types of AI applications heavily rely on publicly available datasets, published not necessarily for that purpose, that they were trained on to build their inner decision-making model, or indiscriminately use as input data to produce new combinations or alterations.¹²⁷ The use of data in such a way might call for not necessarily a new law, but rather a better enforcement of existing laws, namely civil law, protection of author's rights, or intellectual property at large. Alternatively, new mechanisms may need to be introduced, for example that has been happening with DSA¹²⁸ and DMA¹²⁹, although for different reasons, but they share the same legislative zeal and ratio. An example that made it into the legislation, specifically Directive on copyright in the Digital Single Market¹³⁰ that was heavily contested ex post is that search engines need to pay publishers their fair share when using snippets of their articles, likewise the users or creators of AI need to provide a fair remuneration in exchange for using author's data to generate the amalgamated albeit brand new data, that would not have been created if it was not for the authors' original works.

The final aspect I would like to mention is the baked-in human bias and danger of systemic reproduction of incorrect but by observation inferred behaviour of humans into the AI decision-making value hierarchy and logic. There are areas of decision-making which are highly contested for reasons of alleged discrimination, racism or favouritism, namely:

- law enforcement,
- labour market and employee selection/promotion
- financial markets and stock trading,
- public procurement

There is a very real danger that in our desire to objectify decision-making by delegating it to an AI application, we will during implementation unfortunately endow it with baked-in systemic discrimination, which will not manifest initially or too obviously, but nonetheless will be there and will result in distorted decision-making, far from the ideal objective, fact-based, impartial judgement. This usually happens by feeding it with real-world data for training data purposes and accuracy's sake, which has been accumulated over the years by observing the very people, which

¹²⁷ Such as ChatGPT by OpenAI in November 2022, or MidJourney AI by Midjourney, Inc.

¹²⁸ EP – Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act)

¹²⁹ EP – Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act)

¹³⁰ EP – Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC

are meant to be replaced for their inept subjectivity and one-sidedness. The blind faith of big data is a dangerous slipping slope and a Pandora's box. It has been criticised in the past already¹³¹. For that reason, the lawmakers should dedicate attention to not only the AI application itself, but also the training data it is using to pass these judgements.

4.1.2. Cognition (Information and data processing and decision making)

This feature of AI is handling the process of data processing and formulating decisions and answers. First aspect that regulator should handle is that of an AI declaration. End user or recipient of the decision-making process must know that the result he is faced with has been generated by AI, preferably some description of what AI it is, what decision-making architecture it used, and training data, who was the author of the AI. A similar declaration being used in GDPR¹³² legislation, which mandates that user is informed about how his or her data is processed, by whom and for what purpose.

Second there is aspect of cybersecurity and reliability. Should an AI be applied in a high-stake and high-risk situation, the legislator may demand additional guarantees that the robust hardware is used, and that the software is secure. The Cybersecurity Act¹³³ is directly applicable regulation in this regard, here are examples of duties that the cybersecurity act newly imposes:

- continuous risk analysis
- incident reporting,
- drafting of business continuity plans,
- vetting the supply chain security and reliability,
- assuring security in systems acquisition, development and maintenance,
- enforcing security policies and procedures,
- cybersecurity education and training for board members and employees,
- use of cryptography and encryption technologies,
- human resources security, access and asset management,
- use of multi-factor identity authentication, secure communication tools and emergency communication tools

¹³¹ O'NEIL, C. – The era of blind faith in big data must end – youtube.com. [online]. [cit. 2017-09-07]. Available from: https://www.youtube.com/watch?v=_2u_eHHzRto

¹³² EP – Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). eur-lex.europa.eu [online]. [cit. 2023-02-05]. Available from:

<https://eur-lex.europa.eu/legal-content/EN-CS/TXT/?from=CS&uri=CELEX%3A32016R0679>

¹³³ EP – Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act)

The Cybersecurity Act distinguishes two tiers: Essential (very strict) and Important (strict), and out of scope, and allocates certain industries to respective categories depending on their importance for the general security and well-being of critically important infrastructure and services. For example, sectors of energy, transport, banking, healthcare are essential, post services, waste management, food processing, digital services are important. All of these sectors already use AI applications.

Third aspect of data processing is aspect of uncertainty. There is a saying that machines never do mistakes, humans do, and while there is certain wisdom in that quote, for purposes of regulation we need to be prepared for situations, where machines make mistakes also. The question of responsibility and liability for incorrect decision-making made by machine is obvious here and has been exhaustingly debated over the last decade. For further reading I recommend Comparative law study on civil liability for artificial intelligence.¹³⁴ What has not been as debated is the aspect of uncertainty, this is true especially with neural network-based AI, where results are of probabilistic nature. For example, AI is given a picture and is asked to classify what is on said picture.



Picture 3 – AI trying to distinguish a Chihuahua or a blueberry muffin.¹³⁵

While this picture might be entertaining, there is a serious problem at hand here. When an AI decides that certain picture is a muffin, it says so with a calculated probability that reflects its internal training model.

¹³⁴ EC – Comparative law study on civil liability for artificial intelligence.
¹³⁵ SAYRE. M. – The significance of “edge cases” and the cost of imperfection as it pertains to AI adoption papers.ssrn.com. [online]. [cit. 2019-04-25]. Available from: <https://medium.com/@livewithai/the-significance-of-edge-cases-and-the-cost-of-imperfection-as-it-pertains-to-ai-adoption-dc1cebeef72c>

Let us imagine a situation of AI powered pet food feeder, this is what result might look like:

*AI: Based on my internal neural net processing I came to the conclusion that sensor I was equipped with captured an object which contains:
a chihuahua, with 70% probability
a muffin, with 30% probability
therefore, I conclude it is likely chihuahua,
since it is night time, I will dispense food for the chihuahua*

Most of the time, the AI gets it right, some of the time it will not, but the potential damage an automatic AI powered feeder can cause is nearly negligible. Can we say the same for AI powered healthcare? 70% probability of terminal cancer, 30% probability of a 7-day cold. Can we say the same for AI powered military? 70% probability of foe, 30% probability of friend.

It would be advisable to mandate that the probabilities are presented with the result, with explicit warning for the end-user. Much like numbers need to be accompanied by units of measurement, or food products need to be accompanied with information about calories and macros. AI products, their decisions, should be accompanied with metainformation about how really sure the AI is about said result.

The form of the rating for its certainty is also itself an interesting question. Will it be a number, or a percentage? There needs to be defined a unit for certainty. Without such baseline reference the result is meaningless and misleading, which renders the whole legislative initiative pointless. I propose that the result is formulated through two numbers, percentage of matches within dataset (certainty), logarithmic expression of the size of the data set upon which the AI was trained. It needs to be logarithmic to convey sense of degree, which may be very shallow or extremely high. Similarly, how in field of cryptography, a kind of logarithmic approach is used when describing key sizes or a complexity of a computable problem in computer science, or in field of geology, Richter's magnitude scale¹³⁶ for description of earthquake strength.

For example, let us compare two hypothetical example results from a hypothetical AI healthcare provider:

This AI believes you have terminal cancer with certainty of 80% / 10^5 .

This AI believes you have a common cold with certainty of 70% / 10^{16} .

In this extreme example, we have two AI opinions that differ substantially. One predicts patients' death the other a classic seasonal cold. Only thanks to the knowledge of the insufficient

¹³⁶ WILLIAM L. E. "The Richter Scale ML". In Wallace, Robert E. (ed.). [cit. 2008-09-14] The San Andreas Fault System, California. USGS. p. 177. Professional Paper 1515.

scale of the data model of the first AI, we may discard that result and align ourselves with a more robust, though less certain judgement, of the common cold. On first glance the patient can act responsibly and with relevant information. This is a new kind of decision-making people will need to grow accustomed to when dealing with AI applications both in a professional and consumer setting. Obviously, this is an idealistic scenario where the AI is able to correctly self-assess, which is not true and therefore the solution must entail additional measures.

Additionally, these results must be also accompanied with convincing explanation. Explainability of AI application is a hugely important issue. It establishes the underlying framework of trust to the end-users. Between the decision-making agents and subjects receiving the judgement. AI applications might produce convincing enough looking result and accompany that with a robust explanation along with cited sources, which on first glance might seem coherent and reliable, but internally it might be objectively flawed. We describe these moments as AI hallucinations¹³⁷. We cannot take any answers for granted and the explanation for any decision helps us to navigate the thought-process, audit it, double-check it, and potentially troubleshoot it.

A right to explanation is an already established right, that is enshrined both in traditional procedure law, but also specifically hinted at in GDPR, as a representative of a data-related regulation. Though its is debated whether it is binding at all and to what extent does the explanation need to go¹³⁸. Explainability of AI application is a red-line requirement for us to put trust into its decisions, which we will inevitably rely on. As established in the topology of AI, the expert systems can be audited and decision traced and explained step-by-step. However with neural network based AI application that is fundamentally impossible, because there is no way to meaningfully reconstruct the decision-making tree of a multi-billion parameter AI model. Currently we approach AI application as black-boxes, and audit them accordingly based on the relationship between its inputs and outputs. Although there is a growing body of research that challenges the robustness of these mechanisms and questions their reliability¹³⁹.

Another idea might be to mandate built-in “stop and ask” mechanisms, where AI in some applications should have safeguards in place that will automatically make it harmless when met

¹³⁷ "Shaking the foundations: delusions in sequence models for interaction and control". www.deepmind.com.

¹³⁸ WACHTER, S., MITTELSTADT, B., FLORIDI, L., Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation [cit. 2016-12-28]. *International Data Privacy Law*, 2017, Available at SSRN: <https://ssrn.com/abstract=2903469>

¹³⁹ DIMANOV, B., BHATT, U., JAMNIK, M., & WELLER, A. (2020). You shouldn't trust me: Learning models which conceal unfairness from multiple explanation methods. *Frontiers in Artificial Intelligence and Applications: ECAI 2020* <https://doi.org/10.3233/FAIA200380>

with dangerous or uncertain situation, for example self-driving cars, unmanned military grade robots or drones (more on that later). Or applying principles such as “In dubio pro reo”.

Fourth aspect of data processing is of auditability for human review, either for retroactive investigation or subsequent improvement. The decision-making process should ideally be transparent and plainly understandable. Problem is that the inherent quality of neural network architectures that many AIS applications use is that this is not possible.¹⁴⁰ I expect AI audit, AI black boxing and AI stress and QA testing massively impactful and important professions that will soon start to emerge.

Fifth aspect of data processing is the quality of predictive capability, or the ability to process “what if scenarios”. This will predominantly apply for more complicated general AI applications that will assess complex situations, have a lot of data at its disposal and many more options to choose from. The legislator might want to mandate benchmark testing and quality assurance before the AI product may be delivered safely to market. Problem might be “updates” which are often frequent, massive in scope, distributed across the code. It needs to be considered whether it is reasonable to demand an audit for every update when mandating such a requirement.

Sixth aspect of data processing is energy efficiency, though it might be a marginal aspect most of the time. For example, block-chain powered cryptocurrencies are target of recent legislation in China. The legislation outright bans use of crypto in the entire country, mostly for the anti-money laundering concerns but environmental concerns have been mentioned too.¹⁴¹

4.1.3. Agency (ability to affect the world)

The aspect of AI agent’s agency is by far the driving factor behind why we care so much about AI regulation. This aspect describes the scope and scale of how agent can engage and affect the environment. It is the physical manifestation of the AI application in the real world. In chapter Aspects of Epistemology, we discussed the sensors, which served as an input for the AI. In this chapter we will discuss the actuators, which may be considered as the output of the AI. These can range from simple electrical impulses within a wire that represent information, for example “True” or “False”, or they can go all the way to motors, pistons, engines that erect substantial superhuman force and energy, such as robots developed by Boston Dynamics.¹⁴²

¹⁴⁰ SCHMELZER, R. – Towards A More Transparent AI – Forbes [online]. [cit. 2020-05-23]. Available from: <https://www.forbes.com/sites/cognitiveworld/2020/05/23/towards-a-more-transparent-ai/?sh=1792c00e3d93>

¹⁴¹ SAMEK, M., VLASTA, M., – Digital Yuan – Currency or Policy Tool? Acta Universitatis – Carolinae – ISSN: 0323-0619 Karolinum.cz [online]. [cit. 2021-09-13]. Available from: <https://karolinum.cz/en/journal/au-juridica/year-67/issue-3/article-9467>

¹⁴² SZONDY, D. – Boston Dynamics' latest Atlas robot struts its stuff – newatlas.com. [online]. [cit. 2016-02-24]. Available from: <https://newatlas.com/boston-dynamics-new-atlas/42007/>

It would be grave error to say at this point in time that it is inevitable that will create machines that will surpass us in strength and intelligence, because we already have. Humanity has engineered hardware that can go very fast – cars, that can fly really high – planes and rockets, that can crush – hydraulic presses. And this hardware can go to environments previously inaccessible for humans due to its inhospitality, and lasts far longer and is far more energy efficient than any man could be. Let's not limit ourselves to metal and electricity only. Robots can be of biological nature as well. Plant or microbe engineered machines are present-day technology as-well.

I believe we do not need any extra AI specific regulation for Agency aspect of AI applications, instead we should strive for vigilant enforcement of existing legislation, specifically in area of Work & Safety regulation. We should double our efforts into consumer and child protection from dangerous products that may pose a risk. Additionally, we need to become more observant to our ever increasingly designed environments, especially in urban areas, and deploy our hardware in such a way as to minimise unexpected and needless harm.

4.1.4. Autonomy (adaptability, self-governance, self-actualization, self-repair)

First, we need to establish two terms, i.e., what it means for a system to be automated and what it means to be autonomous. When a system is automated, the system is composed of well-defined apriori rules, carefully contained to a very specific scenario with the expectation of exceedingly predictable outcome. When a system is autonomous, the system has capacity for adaptability to changing environment and learns from experience, however still within the bounds of its core programming. Not all AI applications need to be endowed with this feature that lets it reinvent its own programming, either in part, or as a whole. Self-actualisation of inner models of neural network-based AI applications is the core principle behind the technology. It is a method of growth that delivers continuous improvement and innovation through iterative deployments, through series of trials and errors.

There is an interesting subset of problems within this area, some which are of technological but also philosophical nature. An example of technological issue is how to overcome local minima¹⁴³. An example of philosophical issue might be the difference between an alive system capable of evolution and survival, and artificial system designed to mimic alive behaviour.

The technological singularity event¹⁴⁴, described as a moment where an AI application will start to develop new better AI application to replace itself at a pace which is uncontrolled by

¹⁴³ KEIM, R. – Understanding Local Minima in Neural-Network Training – [allaboutcircuits.com](https://www.allaboutcircuits.com/technical-articles/understanding-local-minima-in-neural-network-training/). [online] [cit. 2020-02-06]. Available from: <https://www.allaboutcircuits.com/technical-articles/understanding-local-minima-in-neural-network-training/>

¹⁴⁴ Also known as I. J. Good's intelligence explosion model

humanity, is an object of fascination by many in sci-fi and theoretical computer science literature. That is a certainly an open-ended problem for now, but the pressing problem at hand, which needs to be dealt with right now is unshackled evolution of neural networks that sometimes leads to unexpected and absurd conclusions, which may lead to real-world harm. A case in point that gathered world-wide attention was Microsoft's chatbot Tay¹⁴⁵. In light of this historical experience many future AI developers are taking steps to „neuter“ their AI applications as much as they can to prevent similar backlash in future. The idea behind neutering an AI application is to develop an AI that is capable of specific task, and then overlay that ability with additional layer of ethics checks, which will first evaluate, whether said query or demand by user is admissible. If this additional layer evaluates it is not so, it will refuse to give answer and explain to the user why it refused to execute said action. It is a pre-emptive in-built safety measure. Similar to how a car might start to slow down, or refuse to accelerate, even if the driver wishes it to do so, should the car register an obstacle in front of it with its sensors. The ramifications of machine refusing to do man's bidding are once again very problematic, but it would be naive to expect a machine to obey instruction put forth by human in all circumstances, especially in conflicting instructions, where not even humans can agree on what they want.

I doubt there would be any regulatory framework that would be suitable and sensible to address this specific issue beyond risk-assessment based approaches. Similar problem currently still unresolved exists in cybersecurity legislation. Nature of updates and auditability of updates, either ex ante or ex post is difficult to implement, especially in complex scenarios where many interdependent systems interface between each other.

4.1.5.Purpose (reasons and intent, goals and desires)

Purpose is the why. It is the driving force that precedes all action, the intent, the cause behind the effect. In terms of AI application this aspect deals with reward function, and goal/purpose-oriented AI systems. These reward functions are a measure of how good of a job the AI application is making. The programmer designs the reward function as a means of providing the AI application with a feedback mechanism, to manipulate its actions. In some instances of AI application, the reward function is transparent and objectively measured. For example, in a videogame. There is a score tied to the game world and the AI is given the instruction to maximise its score on the leaderboard, and it will do anything it can to attain that. The reward function is the

¹⁴⁵ KRAFT, A. – Microsoft shuts down AI chatbot after it turned into a Nazi – cbsnews.com [online] [cit. 2016-03-25]. Available from: <https://www.cbsnews.com/news/microsoft-shuts-down-ai-chatbot-after-it-turned-into-racist-nazi/>

mechanism through which goal is defined and measured. In some interested cases of AI application, the reward function is not known to it, and it itself needs to figure out what it might be, based on seemingly random trial and error iterations. It is a scenario where AI systems are trying to understand what the humans want even though the humans are not able to explicitly define that, therefore the AI is limited to observation and inferring the reward function externally.

A real-life example for this are the internet-search algorithms or algorithms that structure the contents of our personalised feeds. The AI system observes what users click on the most, what they engage with the most and tailor future content to maximise the time we direct our attention to it. That is the problem behind addictive nature of social media content, because the AI algorithms successfully figured out how to coerce our brains into addictive patterns and is able to identify content that best responds to each individual's preference. Interestingly, this specific issue has been so important that it has already been addressed by the new Digital Services Act¹⁴⁶. Large platforms that utilise systems which recommend content will need to inform their users how they use these algorithms and also will need to provide their user with a choice, either continue using personalised feed, or offer alternative selection of content that is not based on profiling. That is incredible step forward for AI legislation that is found seemingly in unrelated piece of legislation that deals with Digital Single Market.

We have discussed some of related topics already in previous chapters of Aspects of Autonomy (iterated evolution), and also in Aspects of Aspects of Epistemology (faith in big data). But what is of significance here for the legislator is the opportunity to establish limits for application of AI systems in certain situations. Which intents and goals can be considered lawful, justified and legitimate? Similar question is asked with GDPR, which data processing can be considered as to fall within scope of reasonable and proportional use. Again, we can take a very good inspiration from GDPR which managed to enumerate very good and meaningful categories that establish the legal grounds and basis for personal data processing:

- Consent
- Legitimate interest
- Public interest
- Contractual necessity
- Legal obligation
- Vital interest

¹⁴⁶ EP – Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) Specifically articles 24(a) and 29

I believe this list may be directly application to any AI system and should the application of an AI system fall out of scope of these 6 areas, then it should be deemed as unlawful and illegitimate use of AI application.

Moreover, any intent by the agent must be subject to higher-level instruction, that is in line with the regulatory perspective, there needs to be authentic understanding of the values behind the legislation and embedding that in the code. Compliance on paper, but not in action, and only to the letter of the law, but not the spirit is insufficient. Moreover, there are cases where we might be hard pressed even as human lawyers to give a definitive generalised answer in application of law, and it would be unfair to demand of the machines, i.e., AI applications and the developers of said applications to do the same. Applying the law requires a degree of common sense and interpretation of the regulatory intent. It also requires flexibility and adaptability to resolve a scenario that is completely new and not encountered before. Fundamentally, it is impossible to rule out a need for a new judgement that overrules certain precedent cases. No data model, no matter how sophisticated or how much processing power it would be endowed with can enumerate all possible permutations of combinations and nuances. Application of law is fundamentally about applying two principles that are in conflict and we have to asses in which instances give priority to one over another ad hoc.

The AI would require for these judgements an excellent insight into regulatory truth. Actively obeying the law requires accountability of the agent, which is a concept our legal system is not prepared for yet. It would be more accurate to say the AI application, as a tool, is safe for use according to design specifications laid out by law, or set of design principles laid out by law.

Dr. Svenja Behrendt agrees with such assessment and also finds this aspect very problematic¹⁴⁷. The only meaningful alignment method is to refer the AI to a body of already existing judgements and extrapolating regulatory truths from that and have the AI apply cases to situations it is presented with as best as it can, but monitor each output and allow for correction mechanism and meaningful procedure of human override by default.

Industry leader Open AI takes the alignment issue very seriously and continuously refines their methods and purposefully limits the abilities of their own AI system (designed lobotomy) to bring it closer in line with overall compliance even at the cost of sometimes refusing to obey the

¹⁴⁷ DR. BEHRENDT, S. "Mission impossible? Teaching AI to obey the law", Law and Artificial Intelligence – Challenges and Opportunities, Charles University, Prague, 24. 3. 2023

command of the user of the AI system. The lengths to which these preventive systems might go are concerning and there needs to be more cross-disciplinary research into this area.¹⁴⁸

In order to properly regulate AI, we need to establish a common taxonomy and definition for AI between the fields of Technology and Science and Policy makers.

Certain aspects or qualities such as Epistemic (knowledge of the world), Cognition (the ability to process data), Agency (ability to affect the world), Autonomy (adaptability, self-governance self-actualization, self-repair), Purpose (reasons and desires driving its intent), must be present to a substantive non-trivial degree in order to deem the entity as AI. Due to its non-binary nature, we must work with a spectrum of AI (that may be multi-dimensional due to multiple parameters) that exists within the proposed spectrum of existence. Regulation must recognise these fine details in order to be purposeful and successful in its application.

5. Arising new AI-specific rights and duties

Some of the demands for features arising from the values listed in earlier chapter read very much like a unrealistic wish-list that is unfortunately very broad and open-ended. Engineers and AI data scientists might have a real problem in understanding what is demanded of them, and even a bigger problem in finding meaningful ways how to feasibly implement them. Compliance officers might have troubles identifying what constitutes an active breach of the law or what is an single excess occurrence, or whether AI system is genuinely compliance or merely faking it in a smart way.

The new technology disrupts the balance that current legal system tries to maintain. Therefore, it would be highly advisable to consider whether it is required to establish brand new rights and duties to try to adjust to the new challenges or whether we can make do with what have currently have at our disposal. It could be argued that some new rights in regards to AI are grounded in existing legislation and taking on a brand new meaning within the AI context, therefore they are a mere extension of an existing rule. Although, in some instances that might be pushing it too far and people who would like to excersize those rights might find themselves in precarious and weakened position due to the lack of established present in application. Therefore, we might discover that there is a real need for explicit designation of a brand new right in certain cases, because it is not feasible the legislator had in mind the concept of AI when drafting up legislation.

¹⁴⁸ LEIKE, J., SCHULMAN, J., WU, J., Our approach to alignment research, OpenAI, openai.com [online] [cit. 2023-03-24]. Available from: <https://openai.com/blog/our-approach-to-alignment-research>

Theoretically, we can categorise AI-related rights and duties into two following categories depending on their source of legislation:

- Fundamental – these can be found and derived from charters and constitutional laws and have multidimensional field applicability (generations of human rights, ect.)
- Specific – these can be found in laws that relate to specific issues that usually target area of law (civil, criminal, procedural, ect.)

Further classification of rights can be extrapolated as follows:

- Personal rights (personal integrity, body, appearance, voice, reputation, scoring)
- Free speech related (thoughts, opinions, sharing information)
- Privacy related (confidentiality, use of data)
- Due process (fair trial and procedure, oversight, appeal)
- Intellectual property (authorship, authenticity)
- Civic rights (politics, governance, neutrality)
- Civil rights (liability, remuneration, work-related issues)
- Product safety (standards, specifications, cybersecurity)
- Others

What possible new AI rights and duties might we expect, current legislation that is in effect is explored, as well as its practical application in the real world with provided examples of case law by Court of Justice of the EU within its case law where applicable. For example we have already established AI specific case-law in regards to Intellectual property in other jurisdictions.

For example the United States Court of Appeals for the Federal Circuit has ruled in 2022¹⁴⁹ that AI cannot qualify as inventor for the purposes of filling patents. However, there was a surprising judgement by Federal Court of Australia that in fact for the purposes of patterns and inventions, an AI system can be in fact recognised as the inventor.¹⁵⁰

The list below is by no means exhausting, it merely illustrates few examples of brand new rights and duties which we might find very useful in regards to the use of AI systems that I consider to be of importance going forward. In some cases we will explore whether these rights and duties are grounded in existing or proposed legislation. It should also be noted that any right is closely colerated with a duty of another.

¹⁴⁹ United States Court of Appeals for the Federal Circuit, 2021-2347 Stephen Thaler v Katherine K. Vidal, under Secretary of Commerce for Intellectual Property And Director of the United States Patent and Trademark Office, United States Patent and Trademark Office, [2022] No. 1:20-cv-00903-LMB-TCB

¹⁵⁰ Judgment of the Federal Court of Australia, VID- 108 of 2021 Thaler v Commissioner of Patents, [2021] FCA 879

Further research can be made into specifics of each right or duty and further detail its scope, function, and feasibility. Further research can also be made into listing individual rules into a catalogue and categorising them based on various parameters.

5.1. Potencial AI-specific rights

Arising need for new rights closely correlates to a series of very specific AI-enabled breaches into personal sphere of integrity or sovereignty of an individual. Rights in these cases serve as an optional prerogative and a path to remedy for each individual to counter certain unwelcome force that would have prevailed otherwise. I would like to recognise the arguments laid down by Bajgar & Hořenovský for the support of establishing the regulatory framework around negative human rights¹⁵¹. Though I remain convinced that a degree of specific direct rights are essential to meaningful regulation and protection of interests of natural and legal persons when faced with much faster and stronger opponents, i.e. providers of AI systems.

5.1.1. Right to erasure own data from a data model

Right to privacy can take on many forms and this is one of them. It might resemble GDPR's right to be forgotten¹⁵², but it would be preferable if either legislator in a law, or a judge explicitly confirmed that the right to be forgotten covers such use.

“Extra protections should be given to people whose data have been used to train AI models, such as the right to access models; to know where they have originated from, and to whom they are being traded or transmitted; the right to erase themselves from a trained model; and the right to express a wish that the model not be used in the future¹⁵³.”

While it is perfectly feasible to identify individual and his works when drafting the request for erasure, it might be more challenging to find all related data to that person within the training data. The degree of success depends on the quality of labeling of the data in the datamodel. That is certainly a limiting factor and perhaps even an exceedingly heavy burden to impose.

It is a very similar to the so-far unresolved problem of vast databases of personal data that contains data of users who have given consent, and those who have not. The right answer is to discard the entire database and rebuilt it from ground up, but that is a step rarely taken. Instead processors of personal data attempt to protect their databases under one of the lawful basis for

¹⁵¹ HOŘENOVSKÝ, Jan., BAJGAR, Ondrej. - Negative Human Rights as a Basis for Long-term AI Safety and Regulation [online]. [cit. 2023-06-03]. Available from: <https://arxiv.org/abs/2208.14788>

¹⁵² EP – Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).– Article 17

¹⁵³ STOA – The ethics of artificial intelligence: Issues and initiatives. Brussels: Panel for the Future of Science and Technology (STOA), Scientific Foresight Unit, Directorate-General for Parliamentary Research Services (EPRS) of the Secretariat of the European Parliament. ISBN: 978-92-846-5799-5 Available from: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU\(2020\)634452_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf)

processing personal data besides the consent of the individual, we may expect similar attempts here with training data for AI model.

Case-law on this issue has quite developed since the landmark case of *Google Spain v Costeja* in 2014¹⁵⁴. Though the right is not as straight-forward as one might hope¹⁵⁵. The European Court of Justice ruled in 2019¹⁵⁶ that the Right to be Forgotten did not apply in jurisdictions outside the EU.

5.1.1.Right to corrective action in a data model

The right to corrective action might be understood as a modification of the right to erasure, because it is about the right to modification of training data for reasons of correcting false information, fake news, or even information that is deemed dangerous or unlawful. We ought to tread lightly here as this right will inevitably clash with the freedom of speech and protection from censorship but that is a hotly debated topic of its own. Hate speech or offensive speech also lacks specific and predictable definition that is not coherently unified across the European continent.

Not only the AI regulation in its current draft expects a procedure for corrective actions¹⁵⁷, but content moderation and action against illegal content is already in effect in DSA¹⁵⁸.

In case law we can find a very recent precedent from December 2022¹⁵⁹ for this type of corrective request as an extension of GDPR's right to be forgotten. The condition is that the claimant must submit sufficient evidence capable of substantiating his or her request, by which the inaccuracy of the information found is manifested.

5.1.1.Right to explanation

Right to explanation is a broad and so far not very clearly undefined right. We can take inspiration from procedure law, where judge must convincingly explain his reasoning for his judgements when passing a verdict. Alternatively we can take inspiration from administrative proceedings where authority exercising a competence must make decisions based not on whim but a series of parameters within which it exercises administrative discretion.

¹⁵⁴ Judgment of the Court of Justice of the European Union, C-131/12 *Google Spain SL, Google Inc. v Agencia Española de Protección (AEPD), Mario Costeja González*, ECLI:EU:C:2014:317

¹⁵⁵ DENEMARK, Jaroslav. *Právo být zapomenut v kontextu moderního pojetí ochrany osobních údajů*. Praha : Wolters Kluwer. 2021. 978-80-7552-182-8

¹⁵⁶ Judgment of the Court of Justice of the European Union, C-507/17 *Google LLC, successor in law to Google Inc., v Commission nationale de l'informatique et des libertés (CNIL)*, ECLI:EU:C:2019:772

¹⁵⁷ EC – Proposal for a Regulation of the European Parliament and of the Council on Artificial Intelligence. *Eur-lex.europa.eu*. [online]. [cit. 2021-04-21]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>, – Article 21

¹⁵⁸ EP – Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) - Article 9

¹⁵⁹ Judgment of the Court of Justice of the European Union, C-460/20 *Google LLC v TU, RE (CNIL)*, ECLI:EU:C:2022:962

Unfortunately it is difficult to point to GDPR in this regard because it mandates only that data subjects receive meaningful but limited information about the logic involved in automated decision-making systems, as we touched on earlier¹⁶⁰.

5.1.2. Right to disclosure

First, we can ascertain that at a minimum, the end user should know that he or she engages with an AI application, that itself might not be obvious and enterprises that deploy AI applications go a long way to make sure to mask their AI application usage and to make it appear as human as possible to free up resources of real humans. We can only imagine current series of robocalls, chatbots, phone answering machines, automated e-mails to be only a mere vanguard of things to come.

Second, when we have disclosed we are dealing with an AI, the next step is to give more detailed explanation of the nature of the AI, its function, capabilities, and method of operation. While the first portion might pass with little or no resistance, this second idea might be heavily contested issues. It is because enterprises will want to protect as much of their know-how as possible to keep their competitive edge. Alternatively, they might try to protect the knowledge of inner workings of the AI application to protect it from exploits, or specific manipulation of its actions. In worst case scenario however, they might try to obfuscate and hide gross negligence and illegal operation of the AI application, and that is a serious a problem. There are reports in the UK that AI use in law enforcement has a serious potential to introduce a systemic bias in its deployment.¹⁶¹ We have already mentioned the danger of unquestioned blind faith in big data.

In its current form the proposed regulation establishes public oversight and direct access to data and documentation on AI systems¹⁶² used in the European market, while explicitly assuring of confidentiality so as to protect IP rights, national security and integrity of criminal or administrative proceedings¹⁶³.

Third, when we have established regulatory and independent expert oversight over the overall functioning of the AI application, we can take it a step further. It is reasonable to want an explanation for a decision or judgement by an AI application in specific instances to the user ad

¹⁶⁰ WACHTER, S., MITTELSTADT, B., FLORIDI, L., Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation [cit. 2016-12-28]. International Data Privacy Law, 2017, Available at SSRN: <https://ssrn.com/abstract=2903469>

¹⁶¹ PAULSON, E., UK Police ward of bias in AI tools, ipro.co.uk [online]. [cit. 2019-09-18]. Available from: <https://www.ipro.co.uk/policy-legislation/34435/uk-police-warn-of-bias-in-ai-tools>

¹⁶² EC – Proposal for a Regulation of the European Parliament and of the Council on Artificial Intelligence. [Eur-lex.europa.eu](http://eur-lex.europa.eu). [online]. [cit. 2021-04-21]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>, – Article 64

¹⁶³ EC – Proposal for a Regulation of the European Parliament and of the Council on Artificial Intelligence. [Eur-lex.europa.eu](http://eur-lex.europa.eu). [online]. [cit. 2021-04-21]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>, – Article 70

hoc. For example citing relevant sources from which the AI inferred the result, detailing its thought process in a way that humans can understand, follow and doublecheck step by step, or giving reasons or factual information that it relies on to produce said result. This is a specific explanation for a specific query. However, similar to the second point, it would be fair to disclose similar overall information on the functioning of the AI application as a whole to the end-user whenever it is asked for. Right to explanation in regards to provision of “meaningful information about the logic involved” can be found in GDPR¹⁶⁴. There are also similar rights grounded in legislation dealing with consumer protection in areas of finance, such as MiFID II, or telecommunications, and so on. Recently, we can also point to the DSA¹⁶⁵ that demands Recommender system transparency, which requires providers of online platforms to: *“set out in their terms and conditions, in plain and intelligible language, the main parameters used in their recommender systems, as well as any options for the recipients of the service to modify or influence those main parameters”*. That is precisely the kind of normative text that would and should be applicable to AI applications.

The judgement of what constitutes plain and intelligible language, and sufficient options to influence main parameters however remains a mystery. That is yet another limit of the legislation and it falls to law enforcement agencies to set out standards based on industry and consumer feedback.

5.1.3. Right to redress and human review

Due to delegation of multiple judgements to the AI applications, which used to be uniquely human domain, it is important to realise that we ought to not compromise the quality of judgements while pursuing efficiency and scalability offered by AI applications. Authorities in positions of power who are endowed with the ability to pass judgement and affect with said judgement someone else must and are held responsible for those judgements. There is also an expectation that not all judgements must be fair, or correct, for to err is human, therefore we have developed a series of appeal and redress mechanisms. We can take inspiration from procedure law, which has very detailed and tailored processes depending on the nature of question that is being dealt with in a proceeding (civil, criminal, administrative, insolvency), classes of specialised subjects each with their own function (prosecution, defense, judge), fall back procedures and alternative solutions to problems (ADR, mediation). AI applications are written by humans and are generalised models, while machine is never wrong unless it is broken, in cases of neural networks trained on training

¹⁶⁴ EP – Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) – Article 13 - 15

¹⁶⁵ EP – Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) – Article 27

data it successfully emulates previous mistakes of human teachers. There is no doubt that these mistakes will find its way into its programming, therefore we need procedures to correct them, but most importantly the ultimate authority at the very end should be a human.

The proposed AI regulation approaches this issue from a product safety perspective and establishes human oversight and details its capabilities in more specific detail.¹⁶⁶ While this detail is very comprehensive and well intended only time will tell how will the Article 14 be interpreted and effectively implemented within the landscape of AI applications. One could raise a concern that instead of effective end-user user-friendly design we might see ineffective measures being implemented only to satisfy the bare minimum of legal compliance without truly achieving its intended purpose.

In this chapter we have explored a basic categorisation of new AI related rights and duties. **Fundamental and specific AI rights** differ in their origin and their application, while the former can be found in constitutional-level charters or laws, the latter is found within specific laws that deal with concrete areas of regulation. We have also **given examples of brand-new AI rights and described their use**, and analysed potential inspiration sources from existing legislation or identified proposed implementation of thereof in upcoming legislation. We have also identified other **possible venues for further research**.

¹⁶⁶ EC – Proposal for a Regulation of the European Parliament and of the Council on Artificial Intelligence. Eur-lex.europa.eu. [online]. [cit. 2021-04-21]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>, – Article 14

6. Conclusion

The advent of artificial intelligence is a highly disruptive force that promises to bring about the next industrial revolution, after steam, electricity and internet. So far, a wide range of applications have shown to hold great promise. The benefits of adoption within our society are undeniable. For each opportunity a corresponding risk may be identified. There are many paths to take and we are observers and actors of the next space race of 21st century. From a global perspective, we may expect to observe the unfolding of three distinct approaches, each driven by a unique civilisational undertaking – EU, USA and China. Each of these AI-driven nexuses is built upon a different set of core values, which will undoubtedly find their way into the design and function of artificial intelligence applications.

The phenomena of globalisation changed the world forever and caused it to be more interconnected and interdependent – both a strength and a weakness in its own right. The converging process has seemingly reached its limits. Some might attribute it to the interruption caused by a covid-19 pandemic, others might point to the on-going geopolitical decoupling of the great powers and their ambitions. While the internet was a technology that built many bridges through form of undersea cables and united humanity, the technology surrounding AI might bring about the opposite effect.

One can point to social media that already use some form of AI in their inner functioning and observe the new challenges that are a direct result of it – echo chambers, social bubbles, polarisation, algorithmic manipulation, personalised ad-targeting based on all sorts of criteria directly relating to individual persons, content moderation, protection of copyright, fake news, deep fakes. A broad adoption of AI will only exacerbate these problems further.

The artificial intelligence is also another milestone in the evolution of media and language. Before the printing press, the most effective means of communication was spoken word, bound by geographical limitations, restricted by regional language barriers, and very low speed because of biological limits of the human body.

After the printing press, written word could be delivered and spread on scale. Universal basic education and rise of literacy was the required condition. Suddenly vast swaths of the population could access an up to date information. New forms of government were enabled, one that relied not on feudal management founded on the premise of granting authority to individual to exercise power as he or she sees fit, but a set of generalised rules and guidelines to be interpreted and applied universally across the jurisdiction. Economies of scale allowed cheaper access to information.

The next paradigm shift was the internet, which changed 2 major things. Not only we were no longer bound by distance, as internet is instant irrespective of geographical location. But also changed the fundamental structure of communication. The unidirectional asynchronous communication between an author to a reader was transformed into a synchronous two-way exchange of information, where both people engaged as authors. Not only that, a multilateral exchange of information was enabled, allowing us to reach consensus ever faster on scale, anywhere, though we are still bound by our biological limits, waking hours, attention spans, speed of speech, writing, or thought.

And finally the inconceivable future has arrived, where the artificial intelligence is breaching the biological limits, and speed of thought has been replaced with speed of processing power, which can be scaled exponentially. The implications of attaining this ultimate frontier are unknown. This gap in knowledge and sense of lost control is behind the urgent calls for regulative action. Society is split down the middle in regards to reactions to the advent of artificial intelligence. It can be anywhere between euphoric excitement to existential crisis. This energy is what this paper tried to capture.

The aforementioned context produced the goal of this paper, which was to evaluate methodological approaches to regulation of artificial intelligence. First, the paper asked what were the values that are driving the legislative intent and whether or not the regulation – whatever its form – will be in effect in time. Second, the paper set as its mission to explore what method of regulation was used for the Regulation on Artificial Intelligence, and what type of new responsibilities and law enforcement mechanisms it aims to introduce. Finally, the paper wished to give some examples of new AI-related rights and duties. The goals and methods of research used were described in the first chapter. Then, these research questions were answered in subsequent chapters.

In chapter 2, where the factors and motivations behind why society might want to regulate AI were hinted at. The paper came to conclusion that the purpose of regulation is not really the regulation of the machines themselves, but rather their activity and outcomes, which is decision-making. The paper compared the decision-making of artificial systems to decision-making of humans and predicted that upcoming regulation will have a form of audit-like „thought“ regulation that will be applied to any intelligent system or being capable of decision-making. The paper pointed out the change of narrative, where AI was first viewed as an opportunity, but recently more as a threat.

In chapter 3, the paper has identified a series of documents from various stakeholders that propose their own priorities regarding AI regulation. The values listed in Ethics Guidelines for

Trustworthy AI by AI HLEG were highlighted as they are the most relevant for the regulatory ballance that found its way into the proposed European Regulation on AI. Finally, the paper has listed examples of how human-centric approach to AI regulation manifests in the proposed legislation. Irrespective of the specifics of the letter of the law or the knowledge thereof, the values represent foundational almost constitutional-level ethics on par with human rights. In conclusion of this chapter the paper has utilised the methodological framework set out by N. Petit and explored the five Regulation Models for AI, they are the Black Letter Model, Emergent Model, Ethical Model, Risk Regulation Model and Fifth Model of Externalities with a Moral Twist. Most of the European regulatory proto-initiatives can be subsumed to one of these types of regulatory methodology and as such this topology is a useful guidance and tool to properly asses the functionality of future legislation. The paper came to conclusion that the proposed European Regulation on AI has taken the Risk Regulation approach, by focusing its attention on the process, function and use-case of the AI application rather than the nature of the AI architecture in relation to aforementioned AI topology or datamodel archetypes. and assesed that the Regulation has taken a risk-assesment based model with an accomplying compliance procedure. The paper alludes that undeniably a lot of effort and energy was put forth by the legislators to speed up the legislative process as much as possible – though not at a cost to the quality of the regulation itself - the speed of innovation is still nonetheless far too great and even in the most optimistic scenario, I expect the new rules to be in effect in 2026-2027, which is far too late. Though certainly late is better than never.

In chapter 4 the paper tried to establish that in order to properly regulate AI, it is needed to establish a common taxonomy and definition for AI between the fields of Technology and Science and Policy makers. Certain aspects or qualities such as Epistemic (knowledge of the world), Cognition (the ability to process data), Agency (ability to affect the world), Autonomy (adaptability, self-governance, self-actualization, self-repair), Purpose (reasons and desires driving its intent), must be present to a substantive non-trivial degree in order to deem the entity as AI. Due to its non-binary nature, the regulator must work with a spectrum of AI (that may be multi dimensional due to multiple parameters) that exists within the spectrum of existence. Regulation must recognise these fine details in order to be purposeful and successful in its application.

In chapter 5 the paper has explored a basic categorisation of new AI related rights and duties. Fundamental and specific AI rights differ in their origin and their application, while the former can be found in constitutional-level charters or laws, the latter is found within specific laws that deal with concrete areas of regulation. The paper has also given examples of brand-new AI rights and described their use, and analyzed potential inspiration sources from existing legislation

or identified proposed implementation of thereof in upcoming legislation. The paper also identified other possible venues for further research.

In conclusion I believe this paper has completed its ambition in a sufficient manner. It required a cross-field knowledge and analysis of sources beyond the scope of traditional legal paper. It also gave me the reason and which I transformed into a realised opportunity to attend various conferences that deal with topic of AI regulation, engage with other members of academia and industry outside law faculty and gather a lot of knowledge and insight along the way.

It gave me structure and opportunity to learn and in turn lay out various legal, scientific, societal and moral issues that are related to AI applications as well as offer potential proposals, asses them and highlight known limits to said proposals. There is always room for improvement, especially in enhancing the detail covered, however that would be at a cost of succinct expression of ideas. This paper invites future research to follow-up on questions raised and to adress raised concerns. Personally, I look forward to keeping in touch with latest AI legal developments and new challenges and opportunities that bring along with it.

7. List of abbreviations

AI	Artificial Intelligence
AI HLEG	High-Level Expert Group on Artificial Intelligence
JRC	Joint Research Centre
WEF	World Economic Forum
UN	United Nations
CoE	Council of Europe
OECD	Organisation for Economic Co-operation and Development
EC	European Commission
ECJ	Court of Justice of the EU
EP	European Parliament
IMCO	Committee on Internal Market and Consumer Protection
LIBE	Committee on Civil Liberties, Justice and Home Affairs
STOA	Panel for the Future of Science and Technology (European Parliament)
DG CONNECT	The Directorate-General for Communications Networks, Content and Technology
DGA	Data Governance Act
DMA	Digital Markets Act
DSA	Digital Services Act
GDPR	General Data Protection Regulation
MiFID II	Markets in Financial Instruments Directive
NIS2	Network and Information Security 2 Directive
API	Application programming interface

8. Sources

8.1. Legislation

Treaty on the Functioning of the European Union

Charter of Fundamental Rights of the European Union

Universal Declaration of Human Rights

EP – Regulation (EU) 2021/694 of the European Parliament and of the Council of 29 April 2021 establishing the **Digital Europe Programme** and repealing Decision (EU) 2015/2240. eur-lex.europa.eu. [online]. [cit. 2021-05-11]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32021R0694>

EP – Directive (EU) 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the **legal protection of databases**. eur-lex.europa.eu [online]. [cit. 2023-02-05]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:01996L0009-20190606#M1-1>

EP – Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the proceeding of personal data and on the free movement of such data, and repealing Directive 95/46/EC (**General Data Protection Regulation**). eur-lex.europa.eu [online]. [cit. 2023-02-05]. Available from: <https://eur-lex.europa.eu/legal-content/EN-CS/TXT/?from=CS&uri=CELEX%3A32016R0679>

EP – Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (**Data Governance Act**) eur-lex.europa.eu [online]. [cit. 2023-02-05]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022R0868>

EP – Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (**Digital Markets Act**) eur-lex.europa.eu [online]. [cit. 2023-02-05]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R1925>

EP – Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (**Digital Services Act**) (eur-lex.europa.eu [online]. [cit. 2023-02-05]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065>

EP – Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on **copyright and related rights in the Digital Single Market** and amending Directives 96/9/EC and 2001/29/EC (eur-lex.europa.eu [online]. [cit. 2023-02-05]. Available from: <https://eur-lex.europa.eu/eli/dir/2019/790/oj>

EP – Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (**Cybersecurity Act**) (eur-lex.europa.eu [online]. [cit. 2023-02-05]. Available from: <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32019R0881>

EP – Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing

Directive (EU) 2016/1148 (**NIS 2 Directive**) (eur-lex.europa.eu [online]. [cit. 2023-03-24]. Available from: <https://eur-lex.europa.eu/eli/dir/2022/2555>

EP – European **Declaration on Digital Rights and Principles** for the Digital Decade eur-lex.europa.eu [online]. [cit. 2022-01-26]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A28%3AFIN>

8.2. Proposed legislation

EP – Proposal for a Regulation of the of the European Parliament and of the Council on harmonised rules on fair access to and use of data (**Data Act**) eur-lex.europa.eu [online]. [cit. 2023-02-05]. Available from: <https://eur-lex.europa.eu/legal-content/en/ALL/?uri=COM:2022:68:FIN>

EP – Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual **civil liability rules to artificial intelligence** (AI Liability Directive) eur-lex.europa.eu [online]. [cit. 2022-09-28]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0496>

EP – Proposal for a Directive of the European Parliament and of the Council on **liability for defective products**, eur-lex.europa.eu [online]. [cit. 2023-02-05]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0495>

EC – Proposal for a Regulation of the European Parliament and of the Council on **machinery products**. Ec.europa.eu. [online]. [cit. 2021-04-21]. Available from: <https://ec.europa.eu/docsroom/documents/45508>

EC – **Annex** to the Proposal for a Regulation of the European Parliament and of the Council on **machinery products**. Ec.europa.eu. [online]. [cit. 2021-04-21]. Available from: <https://ec.europa.eu/docsroom/documents/45508>

EC – Proposal for Communication from the Commission regarding the **Declaration of Digital Principles** – the ‘European way’ for the digital society. ec.europa.eu [online]. [cit. 2021-06-16]. Available from: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13017-Declaration-of-Digital-Principles-the-%E2%80%98European-way%E2%80%99-for-the-digital-society_en

EC – Proposal for a Regulation of the European Parliament and of the Council on **Artificial Intelligence**. Eur-lex.europa.eu. [online]. [cit. 2021-04-21]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>

EC – **Annex** to the Proposal for a Regulation of the European Parliament and of the Council on **Artificial Intelligence**. Eur-lex.europa.eu. [online]. [cit. 2021-04-21]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>

EP – Press Release AI Act: a step closer to the first rules on Artificial Intelligence. Europarl.europa.eu. [online]. [cit. 2023-05-11]. Available from: <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>

EP – Draft of Compromise Amendments on the Draft Report Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts. Europarl.europa.eu. [online]. [cit. 2023-05-16]. Available from: <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>

8.3. Case Law

Judgment of the Court of Justice of the European Union, C-131/12 Google Spain SL, Google Inc. v Agencia Española de Protección (AEPD), Mario Costeja González, ECLI:EU:C:2014:317

Judgment of the Court of Justice of the European Union, C-507/17 Google LLC, successor in law to Google Inc., v Commission nationale de l'informatique et des libertés (CNIL), ECLI:EU:C:2019:772

Judgment of the Court of Justice of the European Union, C-460/20 Google LLC v TU, RE CNIL), ECLI:EU:C:2022:962

United States Court of Appeals for the Federal Circuit, 2021-2347 Stephen Thaler v Katherine K. Vidal, under Secretary of Commerce for Intellectual Property And Director of the United States Patent and Trademark Office, United States Patent and Trademark Office,, [2022] No. 1:20-cv-00903-LMB-TCB

Judgment of the Federal Court of Australia, VID- 108 of 2021 Thaler v Commissioner of Patents, [2021] FCA 879

8.4. Literature sources and academic papers

TOMÁŠEK, Michal.; TÝČ, Vladimír a kol.. Právo Evropské unie. 2. vyd.. Praha : Leges. 2017. 496 s.. 78-80-7502-184-7.

SVOBODA, Pavel. Úvod do evropského práva. 5 vyd. Praha : C. H. Beck. 2013.

TOMÁŠEK, Michal a kol. Příklady z evropského práva. 6. vyd. Praha : Ediční středisko PF UK. 2016. 144 s. 978-80-8797-548-0.

DENEMARK, Jaroslav. Právo být zapomenut v kontextu moderního pojetí ochrany osobních údajů. Praha : Wolters Kluwer. 2021. 978-80-7552-182-8

POLČÁK, Radim a kol. Právo informačních technologií [E-kniha]. Praha: Wolters Kluwer, 2018. ISBN 978-80-7598-045-8.

SKULOVÁ, S. Správní uvážení. Brno: Masaryková univerzita Brno., 2003. ISBN 80-210-3237-5

WINTEROVÁ, A., MACKOVÁ, A. a kol. Civilní právo procesní. První část - Řízení nalézací. Vysokoškolská učebnice. 8. nezměněné vydání. Praha: Leges, 2015.

KARNER, Ernst, GEISTFELD, Mark A., KOCH, Bernhard A., Directorate-General for Justice and Consumers (European Commission), ISBN 978-92-76-37819-8, 2021.

VAN ROY, V., ROSSETTI, F., PERSET, K. and GALINDO-ROMERO, L., AI Watch – National strategies on Artificial Intelligence: A European perspective, 2021 edition, EUR 30745 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-39081-7, doi:10.2760/069178, JRC122684.

ROITBLAT, Herbert L. Algorithms Are Not Enough. London: MIT Press. 2020 ISBN: 9780262044127

POLČÁK, R. Internet a proměny práva. 1. vyd. Praha: Auditorium, 2012.

AMODEI D., OLAH C., STEINHARDT J., CHRISTIANO P. Concrete Problems in AI Safety, 2016

- GRACE K., SALVATIER J., DAFOE A., ZHANG B., EVANS O., When Will AI Exceed Human Performance Evidence from AI Experts, 2017
- BROWNSWORD, R., SCOTFORD, E., YEUNG, K. The Oxford Handbook of Law, Regulation and Technology. 1 vyd. Oxford: Oxford University Press, 2017.
- CREMONA, M. New Technologies and EU Law. 1 vyd. Oxford: Oxford University Press, 2017.
- REED, Ch. Computer law: the law and regulation of information technology. 6. vyd. Oxford: Oxford University Press, 2012
- SVANTESSON, D. Private international law and the internet. 1. vyd. Alphen aan den Rijn: Kluwer Law International, 2007.
- SUSSKIND R. The future of law: facing the challenges of information technology. 1. vyd. Oxford: Oxford University Press, 1998.
- OMOHUNDRO S., The Basic AI Drives, 2008
- CHRISTIANO P., LEIKE J., BROWN T., MATRIC M., LEGG S., AMODEI D., Deep Reinforcement Learning From Human Preferences, 2017
- MAY L. (2019). Slavery and Democracy. In Ancient Legal Thought: Equity, Justice, and Humaneness From Hammurabi and the Pharaohs to Justinian and the Talmud (pp. 290-302). Cambridge: Cambridge University Press.
- SWANSON E., & BREWER STEWART, J. (2018). Defending Slavery, Denying Slavery: Rhetorical Strategies of the Contemporary Sex Worker Rights Movement in Historical Context. In E. Swanson & J. Stewart (Eds.), Human Bondage and Abolition: New Histories of Past and Present Slavery (Slavery since Emancipation, pp. 262-293). Cambridge: Cambridge University Press. doi:10.1017/9781316890790.011
- PETIT, N. Law and Regulation of Artificial Intelligence and Robots – Conceptual Framework and Normative Implications. papers.ssrn.com. [online]. [cit. 2017-03-14]. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2931339
- PETIT, N., DE COOMAN, J. – Models of Law and Regulation for AI. Robert Schuman Centre for Advanced Studies Research Paper No. RSCAS 2020/63, EUI Department of Law Research Paper papers.ssrn.com. [online]. [cit. 2020-10-08]. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3706771
- EVARTS, H. Columbia University School of Engineering and Applied Science. AI learns to predict human behavior from videos. Sciencedaily.com [online]. [cit. 2021-06-28]. Available from: <https://www.sciencedaily.com/releases/2021/06/210628113746.htm>
- HEMPHILL, T. A. University of Michigan-Flint. Human Compatible: Artificial Intelligence and the Problem of Control by Stuart Russell. Cato.org [online]. [cit. 2020 Spring/Summer]. Available from: <https://www.cato.org/cato-journal/spring/summer-2020/human-compatible-artificial-intelligence-problem-control-stuart>
- A F SANJUÁN, M. King Juan Carlos University. Artificial Intelligence, Chaos, Prediction and Understanding in Science. [online]. [cit. 2020-03-31]. Available from: https://www.researchgate.net/publication/339711186_Artificial_Intelligence_Chaos_Prediction_and_Understanding_in_Science
- FJELLAND, R. Why general artificial intelligence will not be realized. Humanity Soc Sci Commun 7, 10 [online]. [cit. 2020-06-17]. Available from: <https://doi.org/10.1057/s41599-020-0494-4>

Harvard – Visualisation of Principled Artificial Intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Centre for Internet & Society at Harvard University. cyber.harvard.edu [online]. [cit. 2020-01-15]. Available from: <https://cyber.harvard.edu/publication/2020/principled-ai>

HOŘENOVSKÝ, Jan., BAJGAR, Ondrej. - Negative Human Rights as a Basis for Long-term AI Safety and Regulation [online]. [cit. 2023-06-03]. Available from: <https://arxiv.org/abs/2208.14788>

GREENBAUM, D. (2018). Avoiding Overregulation in the Medical Internet of Things. In I. Cohen, H. Lynch, E. Vayena, & U. Gasser (Eds.), *Big Data, Health Law, and Bioethics* (pp. 129-141). Cambridge: Cambridge University Press. doi:10.1017/9781108147972.013

SHEN, N. AI Regulation in Health Care: How Washington State Can Conquer the New Territory of AI Regulation, 13 *Seattle J. TECH. ENV't & INNOVATION L.* 1 (2023).

KING, S., Why Bigger Is Not Always Better: Dodd-Frank & Its Impact on Small Banks & Businesses, 36 *J. C.R. & ECON. DEV.* 78 (2022).

HARRISON, J., & WOODS, L. (2007). Jurisdiction, forum shopping and the 'race to the bottom'. In *European Broadcasting Law and Policy* (Cambridge Studies in European Law and Policy, pp. 173-193). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511495298.010

KRÁL, Richard. Koncepce a obsah péče řádného hospodáře. Univerzita Karlova. Právnická fakulta, 2016, 74 s.

CIHON, P., KLENALTENKAMP, M. J., SCHUETT, J., BAUM, S. D., AI Certification: Advancing Ethical Practice by Reducing Information Asymmetries, 2021, *IEEE Transactions on Technology and Society*, doi:10.1109/TTS.2021.3077595

SAMEK, M., VLASTA, M., – Digital Yuan – Currency or Policy Tool? *Acta Universitatis – Carolinae – ISSN: 0323-0619 Karolinum.cz* [online]. [cit. 2021-09-13]. Available from: <https://karolinum.cz/en/journal/auc-iuridica/year-67/issue-3/article-9467>

WACHTER, S., MITTELSTADT, B., FLORIDI, L., Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation [cit. 2016-12-28]. *International Data Privacy Law*, 2017, Available at SSRN: <https://ssrn.com/abstract=2903469>

DIMANOV, B., BHATT, U., JAMNIK, M., & WELLER, A. (2020). You shouldn't trust me: Learning models which conceal unfairness from multiple explanation methods. *Frontiers in Artificial Intelligence and Applications: ECAI 2020* <https://doi.org/10.3233/FAIA200380>

KEIM, R. – Understanding Local Minima in Neural-Network Training – *allaboutcircuits.com*. [online] [cit. 2020-02-06]. Available from: <https://www.allaboutcircuits.com/technical-articles/understanding-local-minima-in-neural-network-training/>

WILLIAM L. E. "The Richter Scale ML". In Wallace, Robert E. (ed.). [cit. 2008-09-14] *The San Andreas Fault System, California*. USGS. p. 177. Professional Paper 1515.

8.5. Official publications and advisory documents

AI HLEG. *Digital-strategy.ec.europa.eu* [online]. [cit. 2021-09-17]. Available from: <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai>

AI HLEG – Ethics guidelines for trustworthy AI. *Digital-strategy.ec.europa.eu* [online]. [cit. 2019-04-08]. Available from: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

AI HLEG – Policy and investment recommendations for trustworthy Artificial Intelligence. Digital-strategy.ec.europa.eu. [online]. [cit. 2019-06-26]. Available from: <https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence>

AI HLEG – Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. Digital-strategy.ec.europa.eu. [online]. [cit. 2020-07-17]. Available from: <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>

AI HLEG – Sectoral Considerations on Policy and Investment Recommendations for Trustworthy AI. Futurium.ec.europa.eu. [online]. [cit. 2020-07-23]. Available from: <https://futurium.ec.europa.eu/en/european-ai-alliance/document/ai-hleg-sectoral-considerations-policy-and-investment-recommendations-trustworthy-ai>

EC – Communication: Building Trust in Human Centric Artificial Intelligence. Digital-strategy.ec.europa.eu. [online]. [cit. 2019-04-08]. Available from: <https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence>

EC – White Paper on Artificial Intelligence: a European approach to excellence and trust. Ec.europa.eu. [online]. [cit. 2020-02-19]. Available from: https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

EC – Coordinated Plan on Artificial Intelligence 2021 Review. Digital-strategy.europa.eu. [online]. [cit. 2021-04-21]. Available from: <https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review>

EC – Communication on Fostering a European approach to Artificial Intelligence. Digital-strategy.europa.eu. [online]. [cit. 2021-04-21]. Available from: <https://digital-strategy.ec.europa.eu/en/library/communication-fostering-european-approach-artificial-intelligence>

STOA – The ethics of artificial intelligence: Issues and initiatives. Brussels: Panel for the Future of Science and Technology (STOA), Scientific Foresight Unit, Directorate-General for Parliamentary Research Services (EPRS) of the Secretariat of the European Parliament. ISBN: 978-92-846-5799-5 Available from: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU\(2020\)634452_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf)

CoE – Responsibility and AI. Council of Europe. Expert Committee on human rights dimensions of automated data processing and different forms of artificial intelligence. Rm.coe.int [online]. [cit. 2019]. Available from: <https://rm.coe.int/responsability-and-ai-en/168097d9c5>

OECD – Recommendation of the Council on Artificial Intelligence. Legalinstruments.oecd.org. [online]. [cit. 2019-05-22]. Available from: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

EC – Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions "Towards a common European data space". eur-lex.europa.eu. [online]. [cit. 2018-04-25]. Available from: <https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A52018DC0232>

EC – Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Artificial Intelligence for Europe eur-lex.europa.eu. [online]. [cit. 2018-04-25]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN>

EC – Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - 2030 Digital Compass: the European way for the Digital Decade (Digital Compass policy) eur-lex.europa.eu [online]. [cit. 2021-03-09]. Available from: <https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52021DC0118>

EC/OECD – New report looks at AI national strategies progress and future steps publications.jrc.ec.europa.eu [online]. [cit. 2021-09-16]. Available from: <https://publications.jrc.ec.europa.eu/repository/handle/JRC122684>

EC – Comparative law study on civil liability for artificial intelligence. Op.europa.eu [online]. [cit. 2021-09-16]. Available from: <https://op.europa.eu/en/publication-detail/-/publication/8a32ccc3-0f83-11ec-9151-01aa75ed71a1/language-en>

EC – AI Watch. Defining Artificial Intelligence 2.0. publications.jrc.europa.eu [online]. [cit. 2021-10-29]. Available from: <https://publications.jrc.ec.europa.eu/repository/handle/JRC126426>

CZ – National Artificial Intelligence Strategy of the Czech Republic. Ministry of Industry and Trade. Czech Republic (2019). Available from: https://www.mpo.cz/assets/en/guidepost/for-the-media/press-releases/2019/5/NAIS_eng_web.pdf

EC – Study on Safety and Liability Related Aspects of Software. Digital-strategy.europa.eu. [online]. [cit. 2021-06-17]. ISBN: 978-92-76-37381-0 Available from: <https://digital-strategy.ec.europa.eu/en/library/study-safety-and-liability-related-aspects-software>

UN – Report: The Age of Digital Interdependence. Un.org. [online]. [cit. 2019-June]. Available from: <https://www.un.org/en/pdfs/HLP%20on%20Digital%20Cooperation%20Report%20Executive%20Summary%20-%20ENG.pdf>

WEF – White paper: How to Prevent Discriminatory Outcomes in Machine Learning. World Economic Forum – Global Future Council on Human Rights 2016-2018. Weforum.org [online]. [cit. 2018-03-01]. Available from: https://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf

EP - Report with recommendations to the Commission on Civil Law Rules on Robotics (COM/2020/64). europarl.europa.eu [online]. 19. 2. 2020 [cit. 3.6.2023]. Available from: https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html

EC – Commission Staff Working Document Executive Summary of The Impact Assessment Report Accompanying the Proposal for A Regulation of the European Parliament and of the Council, eur-lex.europa.eu. [online]. [cit. 2021-04-21]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021SC0085>

Data Ethics Commission of the Federal Government Federal Ministry of the Interior, Building and Community, Opinion of the Data Ethics Commission. Bmi.bund.de [online]. [cit. 2019-12-01] Available from: <https://www.bmi.bund.de/SharedDocs/downloads/EN/themen/it->

digital-policy/datenethikkommission-abschlussgutachten-lang.pdf?__blob=publicationFile&v=5

Ministry of Industry, Business and Financial Affairs, New seal for IT-security and responsible data use is in its way. Eng.em.dk [online]. [cit. 2019-10-31] Available from: <https://eng.em.dk/news/2019/oktober/new-seal-for-it-security-and-responsible-data-use-is-in-its-way/>

Parliamentary Secretariat for Financial Services, Digital Economy and Innovation, Towards Ethical and Trustworthy AI. Malta.ai [online]. [2019-08-03] Available from: https://malta.ai/wp-content/uploads/2019/08/Malta_Towards_Ethical_and_Trustworthy_AI.pdf

EC – The First European AI Alliance Assembly. Digital-strategy.europa.eu [online]. [cit. 2019-06-26]. Available from: <https://digital-strategy.ec.europa.eu/en/events/first-european-ai-alliance-assembly>

8.6. Other sources

UN – Urgent action needed over artificial intelligence risks to human rights. News.un.org [online]. [cit. 2021-09-15]. Available from: <https://news.un.org/en/story/2021/09/1099972>

Digitaltrends – R. QUIAN, J. Crime-predicting A.I. isn't science fiction. It's about to roll out in India. Digitaltrends.com [online]. [cit. 2018-04-11]. Available from: <https://www.digitaltrends.com/cool-tech/could-ai-based-surveillance-predict-crime-before-it-happens/>

DICKSON, B. Why AI can't solve unknown problems. venturebeat.com [online]. [cit. 2021-04-02]. Available from: <https://venturebeat.com/2021/04/02/why-ai-cant-solve-unknown-problems/>

KOCH, C. Scientific American – Will Machines Ever Become Conscious? scientificamerican.com [online]. [cit. 2019-12-01]. Available from: <https://www.scientificamerican.com/article/will-machines-ever-become-conscious/>

University of Helsinki, Reaktor Education, elementsofai.com

University of Helsinki, ethics-of-ai.mooc.fi

AI alignment domain – Instrumental convergence, arbutal.com

EK – First calls for proposals under the Digital Europe Programme are launched in digital tech and European Digital Innovation Hubs. Digital.strategy.ec.europa.eu [online]. [cit. 2021-11-17] Available from: <https://digital-strategy.ec.europa.eu/en/news/first-calls-proposals-under-digital-europe-programme-are-launched-digital-tech-and-european-digital>

EK – Commission to invest nearly €2 billion in delivering digital advances to business, citizens, and public administrations. Digital-strategy.ec.europa.eu [online]. [cit. 2021-11-10] Available from: <https://digital-strategy.ec.europa.eu/en/news/commission-invest-nearly-eu2-billion-delivering-digital-advances-business-citizens-and-public>

EK – Call for consultation – Civil liability – adapting liability rules to the digital age and artificial intelligence. Ec.europa.eu [online]. [cit. 2021-10-18] Available from: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12979-Civil-liability-adapting-liability-rules-to-the-digital-age-and-artificial-intelligence/public-consultation_en?s=03

EK – Annual report lays out the challenges of protecting fundamental rights in the digital age. Ec.europa.eu [online]. [cit. 2021-12-10] Available from: https://ec.europa.eu/commission/presscorner/detail/en/ip_21_6699

EK – High-level conference to weigh up EU and international work on Artificial Intelligence. Digital.strategy.ec.europa.eu [online]. [cit. 2021-09-13] Available from: <https://digital-strategy.ec.europa.eu/en/news/high-level-conference-weigh-eu-and-international-work-artificial-intelligence>

EK – International outreach for human-centric artificial intelligence initiative. Digital.strategy.ec.europa.eu [online]. [cit. 2021-09-01] Available from: <https://digital-strategy.ec.europa.eu/en/policies/international-outreach-ai>

UN – High-level Panel on Digital Cooperation. Un.org. [online]. [cit. 2021-06-11]. Available from: <https://www.un.org/en/sg-digital-cooperation-panel>

UN – Artificial intelligence could help to boost the provision of healthcare around the world. Unsplash/Michael Dzedzic. news.un.org [online]. [cit. 2021-09-15]. Available from: <https://news.un.org/en/story/2021/09/1099972>

European Parliament, Artificial intelligence: threats and opportunities europarl.europa.eu [online]. [cit. 2020-05-04]. Available from: <https://www.europarl.europa.eu/news/en/headlines/society/20200918STO87404/artificial-intelligence-threats-and-opportunities>

EC – Fit for 55 - European Green Deal. consilium.europa.eu. [online]. [cit. 2023-06-04]. Available from: <https://www.consilium.europa.eu/en/policies/green-deal/fit-for-55-the-eu-plan-for-a-green-transition/>

SCHMELZER, R. – Towards A More Transparent AI – Forbes [online]. [cit. 2020-05-23]. Available from: <https://www.forbes.com/sites/cognitiveworld/2020/05/23/towards-a-more-transparent-ai/?sh=1792c00e3d93>

SZONDY, D. – Boston Dynamics' latest Atlas robot struts its stuff – newatlas.com. [online]. [cit. 2016-02-24]. Available from: <https://newatlas.com/boston-dynamics-new-atlas/42007/>

SAYRE, M. – The significance of “edge cases” and the cost of imperfection as it pertains to AI adoption papers.ssrn.com. [online]. [cit. 2019-04-25]. Available from: <https://medium.com/@livewithai/the-significance-of-edge-cases-and-the-cost-of-imperfection-as-it-pertains-to-ai-adoption-dc1cebeef72c>

O’NEIL, C. – The era of blind faith in big data must end – youtube.com. [online]. [cit. 2017-09-07]. Available from: https://www.youtube.com/watch?v=_2u_eHHzRto

KRAFT, A. – Microsoft shuts down AI chatbot after it turned into a Nazi – cbsnews.com [online] [cit. 2016-03-25]. Available from: <https://www.cbsnews.com/news/microsoft-shuts-down-ai-chatbot-after-it-turned-into-racist-nazi/>

Conference Law and Artificial Intelligence – Challenges and Opportunities from 24th March 2023

9. Abstrakt v českém jazyce – Metodologický rámec pro Evropskou regulaci (umělé) inteligence a její limity

Tato práce nejdříve zhodnocuje 3 různé přístupy k lidstvu v obecném smyslu slova a dochází k závěru, že je velice těžké kvalifikovat kritéria, která činí přirozeno rozeného člověka nebo právnickou osobu, osobou v právním smyslu, kromě toho, že pro to máme silnou intuici, zatímco u systémů využívající prvky umělé inteligence tato intuice chybí. Práce pokračuje tím, že se definuje umělou inteligenci jako funkční a praktický pojem pro legislativce. Dále přichází ke zjištění, že klíčový aspekt, který žene regulativní mašinerii kupředu je právě schopnost umělé inteligence činit rozhodnutí. V další kapitole práce analyzuje sérii různých dokumentů a studií od různých zainteresovaných stran, které navrhují vlastní priority týkající se regulce umělé inteligence. Byly zdůrazněny priority a hodnoty stanovené v dokumentu Ethics Guidelines for Trustworthy AI od pracovní skupiny AI HLEG, protože je nejrelevantnější s ohledem na navrhovanou regulaci v Evropské unii. Klíčové hodnotové pilíře mají společné to, že se všechny týkají člověka a upínají se směrem k němu. Dále se práce představuje modifikaci topologie umělé inteligence a rozebírá dílčí problémy a kde možné navrhuje řešení nebo zdůrazňuje limity. Práce také definuje základní terminologii a definice pro termíny týkající se umělé inteligence tím, že srovnává výsledky praktických zjištění technických věd v oblasti AI a současnou právní teorii platnou v Evropském prostoru. Dále práce rozebírá metodologický model definovaný Nicolasem Petittem a Jeromem De Coomanem a aplikuje jej na současnou regulaci a kvazi-regulativní dokumenty, které slouží jako legislativní rámec pro budoucí platnou legislativu.

Nakonec práce kriticky zhodnocuje zjištěné požadavky ze současných Evropských regulativních a kvazi-regulativních dokumentů a jejich proveditelnost v reálném a právním světě.

Klíčová slova: Právní rámec, Umělá inteligence, Nová legislativa

10. Abstract in English language - Methodological framework for European regulation of (artificial) intelligence and its limits

This paper first assesses three different approaches to the human condition and arrives at the conclusion that it is exceedingly difficult to precisely qualify the criteria that makes a natural born human, or a legal entity, a Person in legal sense, apart from the fact we have a strong intuition for it, whereas that intuition is lacking for AI systems. The paper proceeds to establish a definition for AI system for the regulator, one that is practical and functional. One of the key aspects that was identified was the ability of decision-making of artificial systems in any form is the driving force behind the need for regulation. Furthermore the paper briefly analyses series of documents from various stakeholders that propose their own priorities regarding AI regulation. Furthermore priorities and values laid out in Ethics Guidelines for Trustworthy AI by AI HLEG were highlighted as they are most relevant for the future regulation. Key values revolve around human-centric approach to AI regulation. Next, the paper explores a modification of introduced AI topology and covers various issues as well as proposals for solutions. The paper explores the fundamental terminology and definitions in regards to AI and AI related terms by comparing the findings of the state-of-the-art theory on AI science and findings of current legal theory used on the European continent. Next the paper discusses the methodological model proposed by Nicolas Petit and Jerome De Cooman and applies it on the emerging regulation and quasi-regulative precursors to legislation in the European Union. In the next chapter the paper introduces a series of examples of brand-new AI rights and describes their use and invites further research in this area. Finally, the paper questions the feasibility of demands and regulative approach undertaken by the European Union so far.

Klíčová slova: Legal framework, Artificial intelligence, New legislation