<u>Jean Monnet Network on EU Law Enforcement</u>
<u>Working Paper Series</u>

The European way to AI in EU Research & Innovation projects

Giacomo Delinavelli[*]

Abstract

The development of Artificial Intelligence (AI) systems has featured in the EU research and innovation framework programmes since 2004 with a specific focus on robotics. The main aim of this paper is to explore the concept of European AI as developed through the Horizon program. Specifically, the analysis focuses on the funding scheme's design, in terms of scope and ethics procedures, also by including specific reference to two Horizon projects dedicated to the development of AI systems in the private and public sector, namely AI4PublicPolicy and STAR-AI. Moreover, this paper will relate to the idea of trustworthiness and inclusion of 'EU values' into the design of artificial intelligence systems. The first part of the paper will focus on guidelines for trustworthy AI systems and recommendations for public policy and investments. The second part will try to connect the effort pursued in Horizon projects with the broader ideas underlying European AI.

[*] Giacomo Delinavelli is a Legal and Policy counsel at Arthur's Legal, in Amsterdam, where he works on various EU (co-)funded projects and corporate matters, on issues regarding AI, Data Governance and Cybersecurity. He has worked as university docent at the University of Amsterdam (PPLE) and Bocconi University, and as a manager for legal affairs at The Good Lobby in Brussels. He holds a law degree at Bocconi University and a Research Master's in Information law at the Institute for Information Law (IViR, UvA). | All opinions contained in this paper should be attributed to the author only, and may in no way be atributed to his employer.

The European way to AI in EU Research & Innovation projects

Abstract

The development of Artificial Intelligence (AI) systems has featured in the EU research and innovation framework programmes since 2004 with a specific focus on robotics. The main aim of this paper is to explore the concept of European AI as developed through the Horizon program. Specifically, the analysis focuses on the funding scheme's design, in terms of scope and ethics procedures, also by including specific reference to two Horizon projects dedicated to the development of AI systems in the private and public sector, namely AI4PublicPolicy and STAR-AI. Moreover, this paper will relate to the idea of trustworthiness and inclusion of 'EU values' into the design of artificial intelligence systems. The first part of the paper will focus on guidelines for trustworthy AI systems and recommendations for public policy and investments. The second part will try to connect the effort pursued in Horizon projects with the broader ideas underlying European AI.

Giacomo Delinavelli*

* Giacomo Delinavelli is a Legal and Policy counsel at Arthur's Legal, in Amsterdam, where he works on various EU (co-)funded projects and corporate matters, on issues regarding AI, Data Governance and Cybersecurity. He has worked as university docent at the University of Amsterdam (PPLE) and Bocconi University, and as a manager for legal affairs at The Good Lobby in Brussels. He holds a law degree at Bocconi University and a Research Master's in Information law at the Institute for Information Law (IViR, UvA). | All opinions contained in this paper should be attributed to the author only, and may in no way be attributed to his employer.

Introduction

In its European AI Strategy,[1] the European Commission (hereinafter also simply mentioned as 'Commission') encourages public and private organisations to develop and adopt Artificial Intelligence (AI) systems that can enhance their operations and ensure safety and fundamental rights. In the Commission's view AI systems are seen as a valid means to increase economic prosperity in Europe, and to deal with pivotal societal challenges, such as climate change, migration and enforcement of laws. At the same time, the development of AI systems in Europe seems to be an objective in itself, a train that cannot be missed, despite its intrinsic risks and potential negative impact on certain weaker social groups.

To craft the 'European way to AI', the Commission has increasingly dedicated public funding to research and innovation programs, such as Horizon. In these programs, public and private organisations explore the potential of AI systems for a more efficient management of services, for the creation of public policies based on large amounts of data, as well as a more effective enforcement of their policies. Nevertheless, it remains an open question how to justify the development of AI systems when fundamental rights, such as privacy and personal data protection, human dignity and non-discrimination are clearly at stake.

The main aim of this paper is to explore the concept of European AI as developed through EU research and innovation programs such as Horizon. Specifically, the analysis focuses on the funding scheme's design, in terms of scope and ethics procedures, also by including specific reference to two Horizon projects dedicated to the development of AI systems in the private and public sector, namely AI4PublicPolicy and STAR-AI. Moreover, this paper will relate the development of European AI systems to the concept of trustworthiness and inclusion of 'EU values' into the design of AI systems.

Firstly, the scope and ambition of the Horizon program with regard to the development of AI systems will be discussed. For the EU, research and innovation are engines of economic prosperity, and programs such as Horizon can enhance civic participation for implementing EU policies and ensure a transparent access to public funding. Moreover, two current Horizon projects will be discussed,[2] namely 'AI4PublicPolicy', which aims to use an AI-powered platform for better public policy making, and 'STAR-AI', which aims to develop AI systems to enhance productivity in industrial manufacturing. The focus on these projects further substantiates the idea of AI systems as a solution for societal challenges and improvement of economic conditions.

The second part of this paper will try to connect the effort pursued in Horizon projects with the broader ideas underlying European AI. The concepts of 'trustworthiness' will be analysed in Chapter 3 and the inclusion of 'EU values' in AI systems in Chapter 4. AI systems developed in research and innovation projects should bring a net benefit to the humans using them, thus deserving their trust for the service they provide. Embedding EU values means to extend the benefits to all Europeans. However, this rosy vision underestimates two crucial aspects: the intrinsic and contextual perils of developing AI systems, as well as the role that AI systems may play in augmenting social conflicts in several environments, such as in the workplace or in the democratic life.

---

[1] European Commission, 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. (COM(2018) 237 final, Brussels, 25.4.2018).

[2] The author of this paper is directly involved in these two Horizon2020 projects, as part of his employment at Arthur's Legal B.V., a legal and strategic firm based in Amsterdam. However, all the opinions expressed in this paper do not necessarily represent the view of Arthur's Legal.

I - Chapter 1: Horizon Europe for the European way to AI

As the global race for AI is gathering pace, the European Union has increased its investment in building digital capacity and funding the development of AI systems. With this commitment, the European Union aims to create economic prosperity, as well as reduce the technological gap among European countries and the rest of the world. Public investments in building technological capacity are common among OECD countries,[3] and in the EU they are disbursed through several funds and programs, the most important being the recent Next Generation EU,[4] Digital Europe,[5] and the Framework Programmes for Research and Technological Development (Horizon).[6] The latter, given its magnitude, focus on research and innovation, and stability in time, will be used as point of reference in this paper.

The program "Horizon Europe", which spans through the period 2021-2027 and replaces "Horizon 2020", provides funding for research and innovation activities to various organisations, such as public and private research institutes, enterprises of any size, and public entities other than research institutions, such as municipalities, hospitals and law enforcement entities.[7] Horizon Europe is articulated in three pillars and multiple programs.[8] In particular, this paper focuses on Pillar II "Global Challenges and European Industrial Competitiveness" where the aim is of creating technical implementations for certain European policies.[9]

Through the Horizon program, the European Union promotes two fundamental values, namely research and innovation as engine of economic prosperity and civic participation for the democratic implementation of EU policies and transparent access to public funding.[10] The research and innovation actions carried out under the Horizon program aim to "address global challenges, including climate

---

[3] The 2019 OECD Council Recommendation on Artificial Intelligence states that governments "should consider long-term public investment, and encourage private investment, in research and development, including interdisciplinary efforts, to spur innovation in trustworthy AI". Izumi Yamashita et Al.,'Measuring the AI Content of Government-Funded R&D Projects: A Proof of Concept for the OECD Fundstat Initiative', vol 2021/09 (2021) OECD Science, Technology and Industry Working Papers 2021/09 <https://www.oecd-ilibrary.org/science-and-technology/measuring-the-ai-content-of-government-funded-r-d-projects_7b43b038-en> accessed 23 April 2023.

[4] It is explicitly stated in the Next Generation EU program that "artificial intelligence will help us to fight climate change and improve healthcare, transport and education." https://next-generation-eu.europa.eu/index_en

[5] https://digital-strategy.ec.europa.eu/en/activities/artificial-intelligence-digital-programme

[6] A good overview can be found here: Francesca Foffano, Teresa Scantamburlo and Atia Cortés, 'Investing in AI for Social Good: An Analysis of European National Strategies' [2022] AI & SOCIETY <https://link.springer.com/10.1007/s00146-022-01445-8> accessed 10 April 2023.

[7] With many caveats, organisations belonging to non-EU states, such as Switzerland, Israel and recently Ukraine may also participate in Horizon projects, although without receiving funding.

[8] The three Pillars of which Horizon Europe is composed are: (1) 'Excellent Science'; (2) 'Global Challenges and European Industrial Competitiveness' (3) 'Innovative Europe'. For a complete overview of Horizon Europe visit: https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en

[9] "investments will be aligned with strategic priorities such as the European Union's Climate Action, the New Industrial Strategy for Europe, Renovation Wave Strategy, European Strategy for Energy System Integration, European Hydrogen Strategy, Offshore Renewable Energy Strategy and Circular Economy Action Plan. Through these investments, the European Union will contribute to digitally transforming the EU industry and to make the EU climate-neutral by transitioning all economic sectors. This will boost the role of the European Union as a solution provider for the benefit of all and position Europe as a technological and industrial leader in the green transition." European Commission, Horizon Europe, Strategic Plan 2021-2024. 12.

[10] REGULATION (EU) 2021/695 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 28 April 2021 establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013. Recital 2, 4.

change and the United Nations Sustainable Development Goals (SDGs), create jobs, boost economic growth, promote industrial competitiveness and boost the attractiveness of the Union in the field of R&I".[11]

In this context, the development of Artificial Intelligence (AI) systems has featured in the EU research and innovation framework programmes since 2004, with a specific focus on robotics. According to the Commission, investments increased to up to EUR 700 million for 2014-2020, complemented by EUR 2.1 billion of private investments as part of a public-private partnership on robotics.[12]

With a total budget of 95 billion, the Horizon Europe program is expected to prioritize investments in the development of AI systems.[13] Although other countries, such as the USA and China, are investing even more on development of AI systems,[14] the European Union is determined in crafting its own way to AI systems, which should respect human rights, be created by multi-disciplinary teams that are inclusive, innovative and optimise for positive social impact, beyond profit. The European way to AI systems supported by Horizon funding is often connected to solving world's biggest challenges, "from treating chronic diseases or reducing fatality rates in traffic accidents to fighting climate change or anticipating cybersecurity threats".[15] Moreover, the Commission itself points out to the potential of AI systems for law enforcement purposes, for instance using AI systems to check how single market rules are applied.

The impact of this funding has been recently critically discussed in a study from the consultancy agency Eticas.[16] In this report, a certain "techno-solutionist" approach from the Commission is noticed. For Eticas, underlying the concept of techno-solutionism there is the assumption that all technological development is good, and that technological solutions can and will eventually solve intricated social problems.[17] Moreover, Eticas also criticized the "supply lead" – or top-down – approach in research and innovation funding, where the development of specific technologies is commissioned by a central public entity, with little if not marginal involvement of civil society organisations.[18] According to the report, the alleged 'impact' generated by the projects is not clearly measured, highlighting a "concerning lack of meaningful impact indicators and evaluation."[19]

Although these claims are based on legitimate concerns, they seem to miss certain nuances. With regard to the "techno-solutionist" critique, assuming that such an approach is actually the one adopted by the Commission, it should be pointed out that basing the development of (AI) technologies in consideration of societal challenges presents significant benefits. The first is to justify the use of public funding, by responding to societal needs that have been considered as of primary importance from political institutions, such as the European Union. Moreover, the Commission places the development of AI systems within the broader context of "digital transformation" of Europe.[20] As the AI strategy

---

[11] Horizon Europe Regulation, Recital 2.
[12] https://eu-robotics.net/divi_overlay/sparc/
[13] COM(2018) 237 final, Brussels, 25.4.2018, 5.
[14] Finding specific figures for 'investments in AI systems in USA and China' should consider direct foreign investments (DFI) of US funds in Chinese AI start-ups, as well as direct public funding.
[15] COM(2018) 237 final, Brussels, 25.4.2018,1.
[16] Eticas, How public money is shaping the future direction of AI: An analysis of the EU's investment in AI development (March, 2023).
[17] Eticas, 25.
[18] Eticas, 27.
[19] Eticas, 29.
[20] For instance, the definition of an AI system in the draft AI Act is currently under debate. Although in the COM's proposal the focus was on the technical features of an AI system, the current version discussed in EU

focuses also on the creation of a digital environment, where the necessary conditions for AI systems such as, availability of data, availability of technical means, and technological savviness, would bring overall societal benefit. And even though the specific societal need is not "solved", technological solutions, such as AI systems, should be considered as part of a "basket" of solutions to societal issues that require complex answers, including but not limited to technological ones.

Regarding the top-down approach to funding research and innovation, as the Eticas report also points out, the Commission targets the funding to certain area where there is little appeal for private investors, thus creating ground technologies for public or private enterprises to pick up and continue.[21] Moreover, a crucial aspect in any Horizon project regard "stakeholders engagement". For this, each project has a dedicated budget for communication and dissemination activities, where the activities and results should be shared with identified stakeholders. In this regard, a tension on the actual meaning of stakeholders' "engagement" could be revealed. Engagement may vary from providing current and quality information to a selected number of people and organisations, to actively seeking inputs which may influence the project's activities and results. This spectrum of possibilities is solved by each individual project, however, the organisations part of the project (also known as 'beneficiaries') balance stakeholders' engagement with the – primary – need of delivering on the expected outcomes.

II – Chapter 2 The AI systems developed in Horizon projects: the cases of AI4PublicPolicy and STAR-AI

Research and innovation programs, such as the Horizon, are funding the development of digital environments for AI systems to provide solutions to identified societal challenges. Funds are allocated to a group of organisations, that together (as a 'consortium') propose a research and innovation project in response to a certain call for proposals. Once the project is awarded, and a detailed Grant Agreement is concluded, private and public entities from across Europe, and sometimes beyond, start working together to implement what promised in the granted project.

The collaboration between private and public entities, and the deriving cross-fertilisation of their competences and cultures, is a core feature in Horizon projects.[22] The common aim is to develop technical solutions that are tailored to the context in which they will be eventually deployed. During the project, AI systems are tested in pilot sites, which, according to the aim of the projects, may be connected to public infrastructures or private organisations. Pilot sites may offer use cases, (raw) data and point out to the specific political and organisational constraints in which the technical systems are supposed to take place. More generally, pilot sites that provide use cases, do offer concrete exemplification of the social needs at which technical partners, in charge of developing the technology, should provide a solution to. The collaboration between public and private entities is meant as an

---

Parliament leans towards the OECD definition: "Artificial Intelligence (AI) System as a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments." OECD, 'Recommendation of the Council on OECD Legal Instruments Artificial Intelligence' (2022).

[21] Horizon Europe, Art. 7(3): 'The collaborative parts of the Programme shall ensure a balance between lower and higher TRLs, thereby covering the whole value chain.' Horizon Europe, Recital 31: "Grants should not be awarded for actions where activities go above TRL 8."

[22] Narrowly defined, public and private partnerships (PPP) are arrangements that public entities use to procure and run public services, such as transportation, health care, energy. In Horizon projects, all the entities, public or private, do not assume any specific legal obligation towards each other. They share the goals set in the Grant Agreement, and with the spirit of loyal collaboration, and in consideration to the specific task allocated to each organisation, they work together to fulfill what stated in the Grant Agreement.

iterative process, in a multiannual period, which aims at the eventual enrichment for all the parties involved, and at the development of more robust, resilient and ethical AI systems.[23]

The AI4PublicPolicy project is a clear case of public and private partnerships (PPP) to solve recognised societal challenges.[24] In this project five municipalities are actively involved as consortium partners.[25] Each public entity provides use cases, that correspond to actual social needs: from dynamic parking availability to enhancing road safety, from maintenance of public infrastructure, such as water pipelines, to optimisation of energy policy. At the same time, technical partners – which include both private enterprises and public research institutes – work to create an AI technical solution (called VPME) that would support policy makers to make decisions based on data and analytics. The general aim of the project is to provide policy makers and public servants with a broader base of information, thus, to make "better" decisions. The project assumes that better decisions are the ones made on a broader information basis, which enhance civic involvement, as well as that enable a more efficient use of scarce public resources. In this sense, the drafting of use cases, the data processing activities as well as the selected Key Performance Indicators (KPIs) all try to substantiate the concept of "better policy making".

The legal and ethical aspects of this project are analysed in an *ad hoc* work package, as well as included in the project management activities.[26] From a legal perspective, it is important to notice that data transfers, to the extent that include personal data as defined in the GDPR, are processed according to the relevant regulations. This requirement is satisfied by each individual data controller and supported by the consortium legal and ethics expert. More importantly, consortium partners also discuss ethical issues during the Ethics Board meetings. For instance, ethical issues concern the identification and engagement of stakeholders, the balancing of opposing interests as well as duties of accountability, explainability and good governance. [27]

In AI4PublicPolicy, the general aim of "better" public policy making is confronted with the need of unbiased, predictable and responsive public actions, which respect fundamental rights, such as privacy, personal data protection, non-discrimination as well as abide to the fundamental principle of good administration. This balancing of interests, for instance collecting more (personal) data against preserving a high level of privacy, is relevant when it comes to the creation of dataset to be fed into AI systems. In these cases, the project partners in charge of the use cases are meant to determine the scope of the data processing activities, as well as the mitigation measures that seem necessary in consideration of the risks possibly occurring. This balancing occurs with the paramount objective of

---

[23] The importance of interdisciplinary collaboration is highlighted in several instances of the European way to AI. Horizon Europe Regulation, Recital 21: "In order to maximise the impact of the Programme, particular consideration should be given to multidisciplinary, interdisciplinary and transdisciplinary approaches as key elements for major scientific progress." High Level Expert Group on Artificial Intelligence, 'Policy and Investment Recommendations For Trustworthy AI': "The roadmap [for European AI] should include major open disciplinary and interdisciplinary fundamental scientific questions and grand challenges of global relevance, as well as major industrial and societal challenges where AI can be an enabler or a significant part of the solution." 21.

[24] A detailed overview of the AI4PublicPolicy project can be obtained on the official website: https://ai4publicpolicy.eu/ as well as on the European Commission website https://cordis.europa.eu/project/id/101004480

[25] These are the Municipality of Genova, Burgas, Lisbon, Athens and Nicosia.

[26] Disclaimer: the author of this paper also works as a legal and ethics expert in the AI4PublicPolicy project.

[27] It should be noted that despite the fact that part of the work carried out in AI4PublicPolicy receives EU funding, not all the Deliverables are immediately accessible to the public due to the sensitiveness of their content. Some Deliverables are publicly available on the Cordis website: https://cordis.europa.eu/project/id/101004480/results

preserving fundamental rights, while fostering the assumption that feeding more data to an AI system will improve its accuracy, and possibly its effectiveness.

In many cases the goal is to automatise certain processes and provide direct supervision – through a data stream – to a central entity. For instance, in one of the use cases of AI4PublicPolicy, the main aim of the AI system is to anticipate the need for maintenance of public infrastructures. This requires collecting relevant data that may provide insights regarding the (degrading) state of the infrastructure. Moreover, citizens in their end-user's capacity are also encouraged to provide information to the central entity regarding the state of certain infrastructures, which may occur in the form of answers to questionnaires or unsolicited input. Although the intended effect of such use cases is to reduce the distance between citizens and public administration, by creating stable and effective communication channels, the undesired effect could be the creation of a "map of complaints" and opinions for a central authority. Assuming that the information received from the citizens, combined with other inputs such as cadastral registries or other environment sensors, allow the identification of persons connecting with the central authority, their right to privacy and personal data protection should be effectively ensured, also with the application of appropriate technical and organisational measures, such as pseudonymisation, restriction to access the information received and limited retention periods.

As shown in this example, AI systems that are intended to provide benefits to the administration of public goods and enhance the participation of citizens could still present detrimental effects to certain fundamental rights. Therefore, the entity in charge of establishing use cases and data processing should implement adequate mitigation measures, such as pseudonymisation, as well as abide by the principles of purpose limitation, data minimisation and storage limitation. Moreover, implementing technical and organisational safeguards, also in consistency with the principles mentioned above, should also aim at counterbalancing the occurring information asymmetry that an AI system may generate in favour of the controlling entity. In fact, as (personal) data are centrally collected, the entity in charge of the processing may strengthen its controlling position, and possibly abuse its power.

Competing interests and the necessity of implementing adequate safeguards is also a topic in Horizon projects that primarily focus on private organisations, such as the manufacturing company that participate in the STAR-AI project.[28] The aim of this project is to design technological solutions, using AI systems, that may contribute to more efficient industrial manufacturing processes. The idea of a more technologically advanced manufacturing industry to enhance efficiency is part of the "Industry 4.0" concept.[29] This policy aims at realising intelligent manufacturing for Europe, where technologies used in other commercial uses, such as cloud-based services, analytics and automated prediction model, may operate on the manufacturing floor. Hence, in response to the call for proposals, STAR-AI project focuses on "Human-robot" collaborations, on use cases that vary from product quality analysis, robot path prediction and efficient process handling.[30]

---

[28] An overview of the STAR-AI project can be found the website: https://star-ai.eu/ as well as on the Commission Cordis website: https://cordis.europa.eu/project/id/956573

[29] Saurabh Vaidya, Prashant Ambad and Santosh Bhosle, 'Industry 4.0 – A Glimpse' (2018) 20 Procedia Manufacturing 233.

[30] In this project, two manufacturing sites are involved, namely Philips B.V. in Drachten (the Netherlands) and OBER-OLIF in Portugal. Moreover, a third site, namely the manufacturing lab site of the research institute DFKI is also involved in the piloting activities.

Although the general aims of Industry 4.0 have been integrated and expanded in the Industry 5.0 policy,[31] the former remains still current in the European industrial landscape, and surely represents the primary objective for this Horizon2020 funded project. The aim of including more advanced technological solutions in industrial processes is a constant in the history of manufacturing,[32] at least since the first industrial revolution, which was followed by other revolutions that brought on the work floor more complex and sophisticated technologies. In all these phases, the role that people, as workers, play in them has always been a paramount concern of any legal and ethical frameworks. Workers have specific legal protections given their particular status, which in Europe are part of an *ad hoc* set of legislations.

With the advent of Industry 4.0 and workers employed in digitised working environments, where their presence is constantly monitored by means of (visual and body) sensors, and decisions regarding their activities are often automatized,[33] the effectiveness of specific guarantees envisaged in employment law and integrated with general and specific provisions of personal data protection law are put into question.[34] In particular, a few question seem urgent: are the established legal safeguards in the labour context up to challenges posed by AI systems? What are the ethical values that are being pursued in this new manufacturing era, Industry 4.0/5.0? How are these values prioritised and who gets to rank them in the first place?

Although this Horizon project provides possibilities to discuss these themes, this conversation is limited to qualified participants, namely the ones who are already developing the technology or will likely use it. Instead, the ones who will likely be subject to its (possible) detrimental effects, namely the workers – also organised in unions – are seldomly included in the conversation. Moreover, even when certain instances, not exactly aligned with the dominant interests, are highlighted, it remains unclear how they are included in the development of the project.

Datasets used to train AI models may provide insights regarding the behaviour of the people included in a use case. How the data are generated also plays a crucial role. For instance, personal data may be derived directly by the use of a communication app. Citizens or users of certain (public) services may communicate their behaviour to a controlling entity, which will act upon that information. When personal data are communicated to a controlling entity, the terms of the engagement should be clear and adequate safeguards should be taken. Nevertheless, personal information may also be derived from sensors and applications that are not meant, immediately, for personal communications, or specifically targeted at individuals. For instance, the data derived from sensors that monitor road traffic, combined with data monitoring parking availability in a certain area, may reveal information regarding mobility, and other habits of a certain group of people or individuals. Therefore, even when personal data are not directly collected, the scope of the use cases, the data processing activities and

---

[31] European Commission. Directorate General for Research and Innovation., *Industry 5.0, a Transformative Vision for Europe: Governing Systemic Transformations towards a Sustainable Industry.* (Publications Office 2021) <https://data.europa.eu/doi/10.2777/17322> accessed 17 March 2023. European Commission. Directorate General for Research and Innovation., *Industry 5.0: Towards a Sustainable, Human Centric and Resilient European Industry.* (Publications Office 2021) <https://data.europa.eu/doi/10.2777/308407> accessed 17 April 2023.

[32] Vaidya, Ambad and Bhosle (n 29).

[33] Jeremias Adams-Prassl, 'What If Your Boss Was an Algorithm?' Comparative Labor Law & Policy Journal 2019 41(1) 123.

[34] Haleform Abraha, Michael Silberman and Jeremias Adams-Prassl, 'The Need for Employee-Specific Data Protection Law: Potential Lessons from Germany for the EU'.

the urban, physical or environmental context, will determine the invasiveness of an AI system on citizens' rights of privacy and personal data protection.

When AI systems are used to provide insights about complex human behaviours, for instance consumer preferences, citizen engagement or workers' productivity, thus, to enhance transparency regarding individual choices, this requires an extended monitoring activity. This activity is meant to further, and is justified on the basis of, the controlling entity's aims; generally, these may regard enhancing productivity, anticipating administrative needs or complaints, targeted interventions. All legitimate interests that will, nevertheless install a monitoring environment, at detriment of weaker, less informed, social groups, such as workers.

Considering that the needs identified by the Commission in its open calls – and further articulated by the project partners that apply for those funding, are of paramount importance for the successful completion of the project, it should be closely considered, in consideration of the technology to be developed, which social group in given organisation is participating to the project and what interests are being pursued. In case one social group high up in the hierarchy, let's say the top management, will apply for funding that will further strengthen its dominant position, additional parties should be involved too.

III – Chapter 3: The concept of 'trustworthiness' on the European way to AI

For the European Commission, systems of Artificial Intelligence ought to be worthy of trust and be used to satisfy recognised societal challenges.[35] Trustworthiness seems to be the first and last requirement of any AI system developed or deployed in Europe, and the same is valid for systems developed in the EU research and innovation projects. AI systems are worthy of trust also when they bring a net benefit to the humans using them, when no one is harmed and when all our legal and ethical principles are abided. However, this rosy vision underestimates two crucial aspects: the intrinsic and contextual perils of developing AI systems, as well as the role that AI systems may play in augmenting social conflicts in several environments, such as in the workplace or in the democratic life.

The Commission is aware that next to the opportunities that a new technology may bring to "humanity", or better, to a certain social group, specific risks to the rights of other (weaker) social groups may also materialise.[36] This awareness was already shown in its 2018 European strategy for AI, when the need for "an environment of trust and accountability" for the development and use of AI was seen as a condition for adoption of AI systems.[37] The risks initially identified referred to the lack of transparency and understanding of the automated outcomes (i.e. "explainability of AI systems"), as well as on the effects of the decisions made with the support of AI systems, which may generate or increase biases and discrimination, and create security and safety issues.[38]

At this point in time, the 2018 AI Strategy did not provide yet a detailed and nuanced picture of the range and severity of risks that may occur using AI systems. However, a clear commitment on exploring these issues more and developing a conceptual "framework" to deal with them was made. Therefore, in June 2018, the European Commission established an independent High-Level Expert Group on Artificial Intelligence (AI HLEG) to provide a framework to deal with the risks that AI systems may

---

[35] Commission, AI Strategy 2021.
[36] Gianclaudio Malgieri and Jędrzej Niklas, 'Vulnerable Data Subjects' (2020) 37 Computer Law & Security Review 105415.
[37] Commission, AI Strategy 2018, 13.
[38] Ibid, 14.

entail.[39] The work of the AI HLEG produced two main deliverables, one focussing on guidelines for trustworthy AI and another one on recommendations for public policy and investments.[40]

Undue influence on human agency, due to subtle acts of manipulation, deception, herding and conditioning was one of the main risks highlighted in the HLEG framework.[41] This risk materialises, for instance, every time information and content "relevant for you" is shown to users of a digital platform. A well-known example concerns machine-learning algorithms used by online search engines, such as Google Search, [42] whose algorithms are fed with personal data from user's online behaviour, ranking of content according to their "*sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender".[43]*

In this sense, the use of AI systems may provoke – intrinsically, i.e., due to the nature of the technology – risks to privacy and personal data protection. Indeed, as machine-learning algorithms require to be trained on vast amounts of good quality data, this may concern the processing of personal or even sensitive data, which should be processed on an appropriate legal basis and in respect of the general principles of purpose limitation and data minimisation.[44] The availability of good quality data that respect privacy and personal data protection is a crucial conundrum for the trustworthiness of European AI systems.[45] For this, the EU has put forth new legislative frameworks, such as the (draft) Data Act and the Data Governance Act, and broader policy objectives, such as the European Common Data Spaces.[46] At the same time, European Data Protection Authorities are engaging with (foreign) AI providers, to avoid detriments to the rights of privacy and personal data protection.[47] Paradoxically, a well-trained AI system may imply a violation of the norms on personal data protection. [48]

---

[39] Nathalie A Smuha, 'The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence' Computer Law Review International; 2019; Vol. 20; iss. 4; pp. 97 - 106.

[40] High-Level Expert Group on AI (AI HLEG) 'Ethics Guidelines for Trustworthy AI' (8 April 2019). High-Level Expert Group on AI (AI HLEG) Policy And Investment Recommendations For Trustworthy AI (26 June 2019).

[41] AI HLEG, 'Ethics Guidelines for Trustworthy AI', 'Human agency' 16.

[42] Petropoulos, G. (2022) 'The dark side of artificial intelligence: manipulation of human behaviour', *Bruegel Blog*, 2 February https://www.bruegel.org/blog-post/dark-side-artificial-intelligence-manipulation-human-behaviour

[43] Michal Kosinski, David Stillwell and Thore Graepel, 'Private Traits and Attributes Are Predictable from Digital Records of Human Behavior' (2013) 110 Proceedings of the National Academy of Sciences 5802. https://www.pnas.org/doi/10.1073/pnas.1218772110

[44] As appeared in the recent intervention from the Italian Data Protection Authority, developers should still comply with GDPR's requirements, for instance clarifying the legal basis they use for the processing of personal data. Italian Data Protection Authority, ChatGPT: Italian SA to lift temporary limitation if OpenAI implements measures 30 April set as deadline for compliance (12/04/2023). https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9874751?s=09

[45] "Garbage in – garbage out" is a colloquial recognition of poor-quality data entry leading to unreliable data output. See, for instance, how Vidgen & Derczynski have analysed the creation of data sets to train machine learning algorithms to detect abusive language. Bertie Vidgen and Leon Derczynski, 'Directions in Abusive Language Training Data, a Systematic Review: Garbage in, Garbage Out' (2020) 15 PLOS ONE e0243300.

[46] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on harmonised rules on fair access to and use of data (Data Act). REGULATION (EU) 2022/868 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 30 May 2022 European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act). European Commission, Staff Working Document on Common European Data Spaces (Brussels, 23.2.2022) SWD(2022) 45 final.

[47] Toby Sterling, European privacy watchdog creates ChatGPT task force. (Reuters, 14 April 2023).

[48] According to van Bekkum and Borgesious, to mitigate biases machine-learning algorithms should be trained on a diverse set of data, which includes special categories of data, such as race, religion and sex. However, this would require the inclusion of a new exception in the GDPR. At the moment, in front of the general prohibition

As machine-learning models grow in complexity, it is very difficult to understand how a certain outcome has been generated. Algorithms as well as the data used to train them are often covered by trade secrets, thus reducing to a minimum the understandability of those systems. Therefore, the use of AI systems may also pose risks to the transparency of the decision-making process, for instance for job recruiting or tax/subsides related matters.[49] The same observation concerns the datasets used to train the machine. But the lack of transparency does not only concern the technical features of an AI system. It concerns also the scope of the system (i.e., the initial query), as well as transparency of the expected outcomes.

The lack of transparency may well increase the asymmetry of power between the decision maker, who avails himself of more complex tools, and the affected subjects, who struggle to understand the decision-making process. Moreover, the people most affected by the impact of AI systems might also be the ones who have less information and control over it, such as children, persons with disabilities and other groups that, in each specific context, have historically been disadvantaged or are at risk of exclusion. As structural discrimination exists also in the most developed countries,[50] the use of AI systems may augment discrimination and affect the life of minority groups. As shown in this seminal study by Buolamwini and Gebru,[51] AI systems may be prone to perpetuating discrimination due to the existing unbalance of power in society. For instance, the availability of comprehensive datasets may limit the results of optimal solutions only for existing elite groups, as well as include in list of suspects people with certain characteristics, such as leaving in certain zip code, having a foreigner last name or a bumpy credit history.[52]

Another risk of AI systems concerns their intrinsic technical vulnerability. As they are integrated in any aspect of daily life, from industrial production to medical diagnostic, from resource management to law enforcement, the consequences of their malfunctioning should be specifically assessed. As any software, it may be subject to malicious attacks, such as data poisoning and model leakage, and produce material harm on people.

A less observed and controversial aspect is the risk that AI systems pose to the environment.[53] As AI systems require large amounts of computing power for the processing of vast amounts of data, scarce

to the processing of special categories of data, as stated in Art. 9(1), Art. 9(2) provides for a set of limited exceptions.

[49] Algorithm Watch, 'How Dutch activists got an invasive fraud detection algorithm banned' (6 April 2020) https://algorithmwatch.org/en/syri-netherlands-algorithm/

[50] NOS nieuws, Racismecoördinator: 'Structurele discriminatie van moslims bij banken' (6 April 2020) https://nos.nl/artikel/2470362-racismecoordinator-structurele-discriminatie-van-moslims-bij-banken

[51] Alex Najibi, 'Racial Discrimination in Face Recognition Technology' (24 October 2020). https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/ Buolamwini and Gebru's 2018 research concluded that some facial analysis algorithms misclassified Black women nearly 35 percent of the time, while nearly always getting it right for white men. Joy Buolamwini and Timnit Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' Proceedings of Machine Learning Research 81:1–15, 2018 Conference on Fairness, Accountability, and Transparency.

[52] Algorithm Watch [49].

[53] Annette Ekin, 'AI can help us fight climate change. But it has an energy problem, too' (Horizon the EU Research and innovation magazine, 12 September 2019) https://ec.europa.eu/research-and-innovation/en/horizon-magazine/ai-can-help-us-fight-climate-change-it-has-energy-problem-too ; NOS Nieuws, 'Datacenters in Noord-Holland kunnen leiden tot drinkwatertekort' (19 March 2021) https://nos.nl/artikel/2373234-datacenters-in-noord-holland-kunnen-leiden-tot-drinkwatertekort ; Dutch Data Center Association, 'Incorrect figures on water usage in data centers cause unnecessary panic' (23 March 2021) https://www.dutchdatacenters.nl/en/nieuws/water-usage-correct-figures/

resources such as energy and water are consumed at scale. The use of scarce resources should be a primary concern for the development of ethical AI systems that can be considered trustworthy.

More recently the Commission, in the Coordinated Plan on Artificial Intelligence 2021 Review,[54] has again expressed the need to consider fundamental rights when developing AI systems, which is both a regulatory requirements, as well as condition for enhancing trust in the technology.[55] This requirement is also engrained in the development of AI systems in EU research and innovation projects, although its meaning might not be immediately understood.

IV – Chapter 4: AI systems built on EU values

As the European Commission is committed to lead Europe in the AI revolution, without renouncing its values, also in European research and innovation projects, safeguards for fundamental rights are embedded in the project's design.

First of all, the project proposal has to answer to a specific societal challenge identified by the Commission. Secondly, in case an AI system is going to involve the development, deployment and/or use of AI systems, a detailed self-assessment whether that could raise ethical concerns related to human rights and values should be produced, as well as detailed risk mitigation measures should be described. Thirdly, Horizon projects create (by default) an "Ethics requirements" Work Package (WP), where specific ethics issues are addressed. This WP normally is led by the project's coordinator with the support of an ethics expert. Finally, in certain cases, the presence of an Ethics board is also envisaged. The project coordinator will invite relevant project partners and the ethics expert will assist the board's activities. Although these procedural safeguards may ensure that relevant ethical concerns are identified and addressed, the effectiveness of these safeguards will depend on how the relevant parties are able to strike a balance between technical constraints and ethical considerations.

References to European values in relation to the development of AI systems is a constant in several EU policy documents,[56] and further specified in the reviewed European strategy for AI.[57] Given the consistency of this reference, it is worth inquiring about its actual meaning.

European values are generally stated in Art. 2 Treaty on the European Union. This article explicitly includes "human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities. These values are [meant to develop] a society in which pluralism, non-discrimination, tolerance, justice, solidarity and equality between women and men prevail." The observance of these overarching values, the European Court of Justice has considered how intertwined they are to norms that specify their meaning and give concrete implementation.[58] Therefore, to list of "values" that need to be observed when developing AI systems,

---

[54] European Commission, Coordinated Plan on Artificial Intelligence 2021 Review [CITE SPECIFIC PAGES]

[55] COM(2018) 237 final, Brussels, 25.4.2018, 14.

[56] "The European Union has the ambition of empowering European citizens with digital solutions rooted in our common values […]" Horizon Europe, Strategic Plan 2021 – 2024, 7.

[57] "AI should work for people and be a force for good in society'. Given the major social and environmental impacts of AI technologies, a human-centric approach to their development and use, the protection of EU values and fundamental rights such as non-discrimination, privacy and data protection, and the sustainable and efficient use of resources are among the key principles that guide the European approach. European Commission, ANNEXES to the 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 Fostering a European approach to Artificial Intelligence (Brussels, 21.4.2021, COM(2021) 205 final ANNEX).

[58] Judgments of 16 february 2022, Hungary v Parliament and Council, C-156/21, EU:C:2022:97, paragraphs 157– 158; and of 16 february 2022, Poland v Parliament and Council, C-157/21, EU:C:2022:98, paragraphs 193–194

reference should be made to certain provisions of the Charter of Fundamental Rights of the EU (CFREU), as well as to secondary legislation, such as the General Data Protection Regulation (GDPR) and the upcoming AI Act. Moreover, as primary and secondary legislation could still not be able to provide all the answer to face the complexity of AI systems, the EU has developed a High-level Expert Group Ethical Guidelines.[59]

AI systems developed in EU research and innovation projects are meant to improve the individual and collective condition of European citizens, thus the reference to EU values, from a moral perspective, could be intended also from a utilitarian perspective: aggregate benefits should be more than the potential harms. The balancing of benefits and harms can be analysed from different perspectives. [60]

From a citizen-perspective, the development of AI systems means to "empower" European citizens with tools and legal entitlements, thus, allowing citizens to reap the benefits of the digital economy. This empowerment is articulated, through EU research & innovation projects, in at least two ways. The first concerns favouring the development of technical skills for individuals, which would allow European citizens to understand and be a meaningful participant not only of the consumption of technology, but also a critical member of its production. For instance, in STAR-AI a specific AI system is meant to suggest to workers training and education programs that can help them to remain competitive on the job market, or in AI4PublicPolicy, local stakeholders are called to participate in co-creation activities, where the limits and the potential of the AI system (VPME) are discussed to address their own concerns.

Secondly, as European citizens are provided with additional legal entitlements, regarding, for instance the right to control their personal data and non-personal data,[61] these are meant to give them a position of strength vis-á-vis other market participants. In this way, the AI systems developed in EU projects aim to create products and services for making good of the promised benefits.[62]

From an economic perspective,[63] the adoption of AI systems is seen as a periodical "upgrade" of the European industrial system in consideration of the technology available. The different phases of the European industry, also marked as software updates, the latest being Industry 4.0 and Industry 5.0 bring new considerations about the relationship between technology and society. With Industry 4.0, the goal is to prompt all the various components of the industrial production to uptake digital tools already available in the consumer market, such as cloud and edge computing as well as AI systems to enhance productivity and competitiveness on global markets. Whereas the concept of industry 5.0 goes well beyond the technological transformation of the industry. Industry 5.0 is characterised by the shift of focus from generating value exclusively for shareholders, to generating value for all the stakeholders. This implies that industry should keep into high consideration the global challenges that society is facing, in a world of 'polycrisis',[64] and be cognizant of the specific challenges that each stakeholder is facing.

---

[59] *Supra*, 40.
[60] Vanja Skoric, Giovanni Sileno and Sennay Ghebreab, 'Legality, Legitimacy, and Instrumental Possibility in Human and Computational Governance for the Public Sector (IFDaD 2022 INTERNATIONAL FORUM on DIGITAL and DEMOCRACY, November 17-18, 2022, live and digital event in Rome, Italy)
[61] For instance, right to portability of data as envisaged in Article 20 GDPR and Article 5 draft Data Act.
[62] For instance, projects such as FAME and PISTIS aim to deliver a federation of data marketplaces, where citizens' data may be valorised. See: find more info for FAME at https://cordis.europa.eu/project/id/101092639 and PISTIS at: https://cordis.europa.eu/project/id/101093016
[63] Inga Ulnicane, 'Emerging Technology for Economic Competitiveness or Societal Challenges? Framing Purpose in Artificial Intelligence Policy' (2022) 2 Global Public Policy and Governance 326.
[64] Adam Tooze, 'Welcome to the world of the polycrisis' (Financial Times, London) 28 October 2022.

In EU research & innovation projects, the analysis of societal challenges proceeds technological development. Tools and processes are designed with a clear idea in mind of the societal challenge that the project wants to overcome. For instance, the concept of Industry 5.0 recognises the power of industry to deal with societal goals, beyond jobs and growth, while becoming a resilient provider of prosperity, by making sustainable production and placing the wellbeing of the industry worker at the centre of the production process. For the Commission, Industry 5.0 should be based on the values of human-centricity, sustainability and resilience. [65]

From an international perspective, embedding EU values in AI systems means to create the availability and ensure the control, for the European industry and political actors, of the most advanced technologies, thus, to not be subject to "others'" values. This ambition is often encapsulated in the concept of "European Digital Sovereignty".[66] Technological competition has always been a bone of contention among nations. It is easy to think at the race to the space, between the US and the USSR, sparked when the latter sent the first satellite in orbit. Or to the most recent race for the development of the vaccine for the Covid-19 pandemic. Being in control of the most advanced means of production, can mean prominence on the global stage. This is reflected by the urgency for the Commission to "catch up" on AI, or at least, to not lay (too much) behind other global competitors, namely US and China. [67]

Concluding remarks

The European Commission in its 2018 strategy has identified AI systems to be a valuable technology for solving certain societal challenges. For this, research and innovation programs such as Horizon have been funding the development of trustworthy AI systems. These systems present several intrinsic risks, related to their opacity, need of personal and sensitive data and natural resources, such as energy and water. Moreover, the development of trustworthy AI systems should occur on the basis of EU values, such as the respect for human dignity, accountability and digital sovereignty. Embedding EU values in AI systems is an ambition that is also included in European research and innovation such as Horizon, with specific procedures and safeguards. Nevertheless, the effectiveness of these safeguards may be limited by the scope of these projects, which may require balancing technical constraints with ethical considerations. Horizon projects such as AI4PublicPolicy and STAR-AI do have ethical procedures in place to mitigate intrinsic and contextual risks and include EU values in the design and deployment of AI systems.

However, the selection of use cases the aims that AI systems are meant to achieve may augment power asymmetries and reinforce the position of dominant groups, such as the management within a company or a central authority within a municipality.

---

[65]European Commission. Directorate General for Research and Innovation. (n 31). 6.

[66] European Commission, has identified digital policy as one of the key political priorities of her 2019-2024 term in office and pledged that Europe must achieve 'technological sovereignty' in critical areas. Tambiama Madiega, 'Digital Sovereignty for Europe' (European Parliamentary Research Service, July 2020).

[67] (COM(2018) 237 final, Brussels, 25.4.2018), 2

References

Abraha H, Silberman M and Adams-Prassl J, 'The Need for Employee-Specific Data Protection Law: Potential Lessons from Germany for the EU' (European law blog, 30 September 2022) https://europeanlawblog.eu/2022/09/30/the-need-for-employee-specific-data-protection-law-potential-lessons-from-germany-for-the-eu/ accessed on 2 May 2023.

Adams-Prassl J, 'What If Your Boss Was an Algorithm?' [2019] 41(1) Comparative Labor Law & Policy Journal 123.

Buolamwini J and Gebru T, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' [2018] Proceedings of Machine Learning Research 81 1-15.

European Commission. Directorate General for Research and Innovation., *Industry 5.0, a Transformative Vision for Europe: Governing Systemic Transformations towards a Sustainable Industry.* (Publications Office 2021) <https://data.europa.eu/doi/10.2777/17322> accessed 17 March 2023.

——, *Industry 5.0: Towards a Sustainable, Human Centric and Resilient European Industry.* (Publications Office 2021) <https://data.europa.eu/doi/10.2777/308407> accessed 17 April 2023.

——, Directorate General for Research and Innovation, *Horizon Europe: Strategic Plan 2021 2024.* (Publications Office 2021) <https://data.europa.eu/doi/10.2777/083753> accessed 1 March 2023.

——, *Industry 5.0, a Transformative Vision for Europe: Governing Systemic Transformations towards a Sustainable Industry.* (Publications Office 2021) <https://data.europa.eu/doi/10.2777/17322> accessed 17 March 2023.

European Court of Auditors, Public Private Partnerships in the EU: Widespread shortcomings and limited benefits (Special Report, 2018).

European Union Agency for Fundamental Rights, *Getting the Future Right Artificial Intelligence and Fundamental Rights; Report* (Publications Office of the European Union, 2020).

Foffano F, Scantamburlo T and Cortés A, 'Investing in AI for Social Good: An Analysis of European National Strategies' [2022] AI & SOCIETY <https://link.springer.com/10.1007/s00146-022-01445-8> accessed 10 April 2023.

Graber CB, 'Artificial Intelligence, Affordances and Fundamental Rights' in Mireille Hildebrandt and Kieron O'Hara, *Life and the Law in the Era of Data-Driven Agency* (Edward Elgar Publishing 2020) <https://www.elgaronline.com/view/edcoll/9781788971997/9781788971997.00018.xml> accessed 3 March 2023.

High Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI (2019).

High Level Expert Group on Artificial Intelligence, Policy and Investment Recommendations for Trustworthy AI (2019).

Kosinski M, Stillwell D and Graepel T, 'Private Traits and Attributes Are Predictable from Digital Records of Human Behavior' (2013) 110 Proceedings of the National Academy of Sciences 5802.

Malgieri G and Niklas J, 'Vulnerable Data Subjects' (2020) 37 Computer Law & Security Review 105415.

Yamashita, I., et al. (2021), "Measuring the AI content of government-funded R&D projects: A proof of concept for the OECD Fundstat initiative", *OECD Science, Technology and Industry Working Papers*, No. 2021/09, OECD Publishing, Paris, https://doi.org/10.1787/7b43b038-en. accessed 23 April 2023.

Munn L, 'The Uselessness of AI Ethics' (2022) AI and Ethics <https://link.springer.com/10.1007/s43681-022-00209-w> accessed 9 January 2023

NIST, Artificial Intelligence Measurement and Evaluation at the National Institute of Standards and Technology (2021).

Smuha NA, 'The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence' (Computer Law Review International; 2019; Vol. 20; iss. 4; pp. 97 - 106)

Tambiama M, 'Digital Sovereignty for Europe' (Briefing to the European Parliament, 2 July 2020).

Ulnicane I, 'Emerging Technology for Economic Competitiveness or Societal Challenges? Framing Purpose in Artificial Intelligence Policy' (2022) 2 Global Public Policy and Governance 326

Vaidya S, Ambad P and Bhosle S, 'Industry 4.0 – A Glimpse' (2018) 20 Procedia Manufacturing 233

Vidgen B and Derczynski L, 'Directions in Abusive Language Training Data, a Systematic Review: Garbage in, Garbage Out' (2020) 15 PLOS ONE e0243300