Special Committee on Artificial Intelligence in a Digital Age

European Parliament

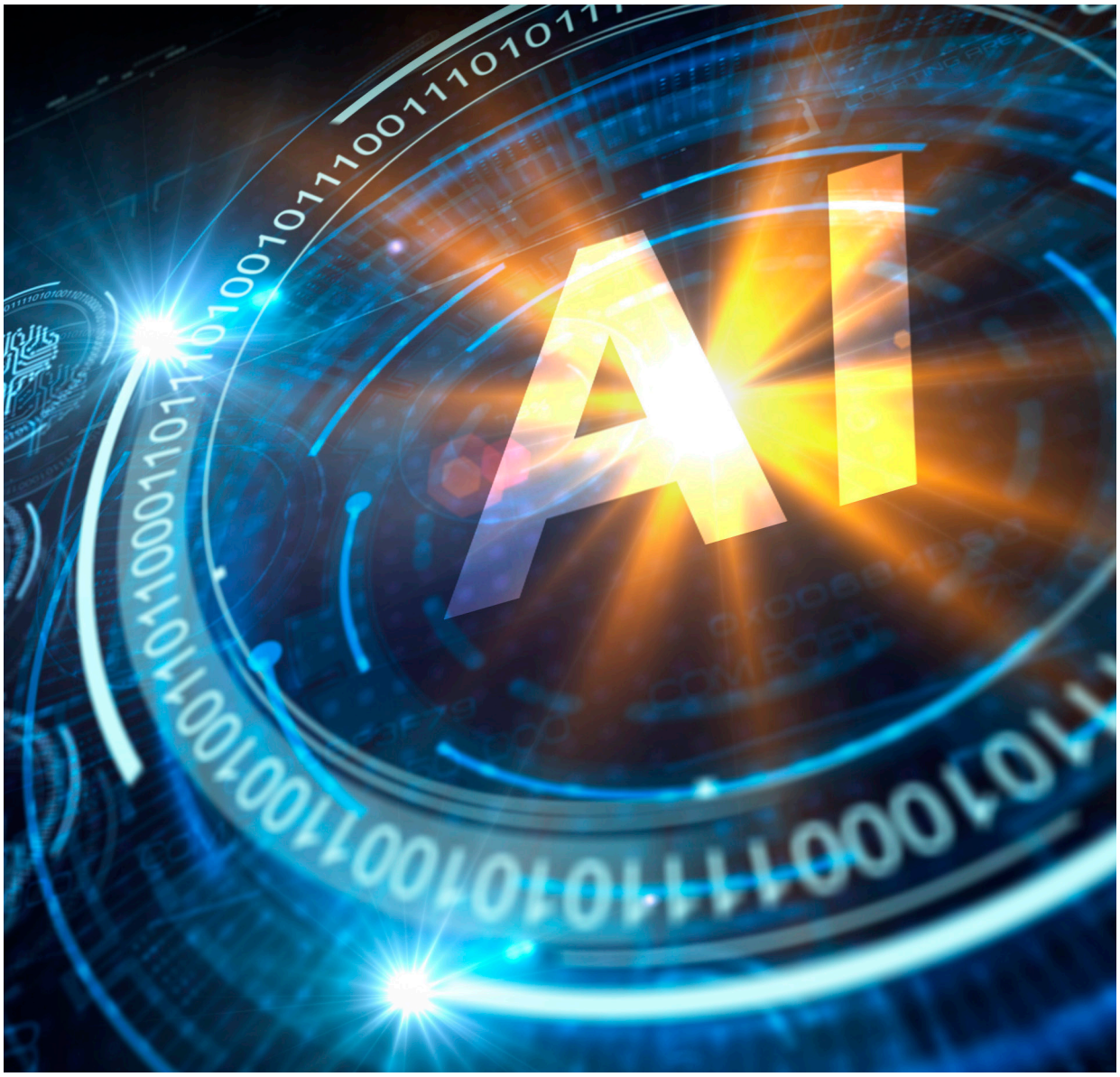
AIDA Working Paper on

AI and the Data Strategy

following the AIDA Public Hearing on 30 September 2021

November 2021 | WWW.EUROPARL.EUROPA.EU - @EP_ARTIFINTEL

Introduction



The special committee on Artificial Intelligence in a Digital Age (AIDA) organised a public hearing on "AI and the Data Strategy" to explore perspectives on the handling of data in the context of AI. The hearing encompassed two panel discussions that brought together experts from industry, consumer organisations, and academia. Both panels presented a variety of challenges and opportunities regarding data sharing, data use and the link thereof to AI development and deployment.

In the AIDA Vice-Chair's opening remarks, Ms. Miapetra Kumpula-Natri highlighted that the three key aspects of AI are skills, infrastructure, and data, while noting that data has often not received enough attention as the necessary third variable. In the Vice-Chair's view, data is the driving force in AI implementation; if the algorithm is the engine, then data is the fuel. For this reason, data must be a central focus of the EU strategy on AI adoption. The Vice-Chair concluded by recalling also the importance of trust, equity, and values and emphasised that these principles will continue to guide the EU's data policy going forward.

Part I: Exchange of views with the European Commission

Intervention:

• **Mr Kilian Gross**, Head of Unit, AI Policy Development and Coordination, DG-CNECT, European Commission

Mr Gross reemphasised the essential preconditions for the EU becoming an AI hub, which are data, computer infrastructure, and skills. Furthermore, Mr Gross outlined key programs and acts that will position Europe as a key player in AI, which are the Digital Decade Policy Programme, the AI White Paper, and the Data Act. The transatlantic partnership is also central to the EU approach to AI, which is exemplified in the EU-US Trade and Technology Council (TTC), Mr Gross noted.

Additionally, Mr Gross outlined key targets of the EU set forth in the Digital Compass, which included training more skilled workers, increased connectivity and 5G in the EU, building the first computer with quantum acceleration, increasing cloud and Big Data usage within EU companies, as well as public services becoming entirely available online. To achieve these targets, an annual cooperation cycle is necessary to develop EU trajectories and monitor progress in which Member States draft individual strategic roadmaps tracking the progress on their commitments. In light of the proposed Digital Governance Act (DGA) [1], the upcoming proposal for a Data Act[2], and the proposed AI Act[3], it is also crucial for the EU to make data a central issue to the AI strategy in the future. Mr Gross concluded by recalling that data has become one of the most powerful resources in the world and is particularly crucial in the development and application of AI.

1 Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act) COM/2020/767 final

2 (Not yet published) - See: Communication from the Commission - A European strategy for data (COM(2020) 66 final)

3 Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts (COM/2021/206 final)

Part II: Exchange of views with the Civil Society:

Key Takeaways - Panel I:

The first panel focused on the perspectives of representatives of industry and civil society and was composed of four experts.

- **Mr Jeremy Rollison,** Senior Director, Data Policy & Digital Inclusive Economy, Microsoft
- **Ms Gemma Galdon,** Founder of Eticas Foundation and Eticas Consulting
- **Mr Stefaan G Verhulst,** Co-Founder and Chief Research and Development Office of the Governance Laboratory (The GovLab) at New York University (NYU)
- **Mr Walter Palmetshofer,** Policy and Research at Open Knowledge Foundation, Germany

Mr Rollison of Microsoft raised concerns about data equity stating that **in the distribution of data there is an increasing risk of concentration of data into the hands of particular countries or companies. Such a "data divide" will eventually reinforce biases within AI technologies.** Mr Rollison also worried that the value of data may be lost when it is not analysed by people with the right expertise and tools and regarded data clustering as a major roadblock for the development of AI technology. To close the data divide, Mr. Rollison suggested that the sharing of data should become a common practice through the use of open data sources which allow for increased access, sharing, and reuse of data. Mr Rollison also mentioned that Microsoft has launched an open data campaign in the pursuit of creating an environment conducive to data-sharing.

Ms Galdon of Eticas Consulting shared her doubts about the general applicability of current AI algorithms and noted that **AI systems developed for the consumer market are increasingly being used in high risk situations. In the opinion of Ms Galdon, the same low-quality algorithms made for marketplaces are being applied to public sector contexts in health, education, security, and social services. This is AI that was designed to optimise for scalability and profit, not compliance and social impact.** Ms Galdon explained that in the present state of the AI market, novel software goes straight from the technology companies to consumers without oversight. Ms Galdon made the point that not all stakeholders are involved in the current market environment and that AI systems are not developed using a multidisciplinary approach. Ms Galdon concluded that the market led by Silicon Valley does not provide sufficient conditions to produce AI for high-risk situations. **For this reason, the EU must go beyond regulatory leadership on AI and promote a market that develops technologies that are sufficiently mature to meet the high-risk challenges we expect technology to solve.**

Mr Verhulst of GovLab outlined that proper data analysis has the potential to make governing more legitimate and promote better policy decisions. However, **there currently exists a massive asymmetry in data access, computational power, and AI expertise, which is leading in many cases to the risk of data misuse. Averting data misuse can be accomplished through a mechanism known as "data collaboratives".** According to Mr Verhulst, "data collaboratives" are systems in which participants from different sectors of the economy - including private companies, research institutions, and government agencies - can exchange data to help solve public problems. **Mr Verhulst also proposed purposeful re-utilisation of data, noting that a stronger determination for the reuse of data will strengthen societal demand for data collaboratives. Additionally, Mr Verhulst recommended more public involvement and support for data collaboratives, which could be accomplished through intermediaries between the collaboratives and citizens.** Participation in the form of data stewards and citizen assemblies could help to decide which data reuse is desirable and which is problematic. Lastly, Mr Verhulst emphasised the need for more data analysis on best practice policies in order to understand which data models work and which do not.

Mr Palmetshofer of the Open Knowledge Foundation focused on the time delay in data acquisition, articulating that **there is still a serious lack of real time data for industries such as health and transportation. To reach the goals set by the EU in the "Digital Decade" the process of acquiring data must be vastly accelerated via the Open Data Directive.** Mr Palmetshofer suggested that the Open Data Directive must be transposed into international law and that a true shift towards open data will require a change in the current closed data culture where data is sold for a profit by businesses. As Mr Palmetshofer stated, civil society should not be "begging companies to share their data". Additionally, Mr Palmetshofer also expressed concerns that Europe needs a more concrete investment in data to accomplish better research and better training of its workforce, otherwise he fears that the "digital decade" will not shape up according to EU requirements.

Key Takeaways - Panel II:

The second panel concentrated on the perspectives from members of industry and civil society regarding data in AI and was composed of four experts.

- **Mr Prof. Luis Paulo Reis,** University of Porto, Portugal
- **Ms Prof. Luisa Specht-Riemenschneider,** Data law specialist, University Professor at the Rheinische Friedrich-Wilhelms-Universität Bonn
- **Mr Thomas Bolander,** PhD, professor of Artificial Intelligence (AI) at DTU Compute, Danish Technical University
- **Ms Sarah Chander,** Senior Policy Advisor at European Digital Rights (EDRI)

Mr Luis Paulo Reis of the University of Porto explained that there is a significant opportunity for AI in the field of public administration. However, to realize these opportunities, Mr Reis suggested that a free flow of data must be established. Mr Reis also stated that **in current applications of AI there must be more emphasis on human oversight and decision-making. To incorporate AI into decision-making processes, we need configurable AI, meaning AI that is regulated for the EU but that is also customizable for a given region, country, and citizen.** According to Mr Reis, effective AI must be tailored for its intended purpose. This will require opening up and allowing for greater free-flow of non-personal public sector data as well as anonymized personal data. Lastly, Mr Reis urged that specific AI funding for public administration should be granted.

Ms Specht-Riemenschneider, a data law specialist and University professor at the Rheinische Friedrich-Wilhelms-Universität Bonn, focused on the concept of data oversight in research recalling that **currently the largest holders of data are big tech companies and if this is to change, then data governance laws must create incentives for the development of data trustees.** Such data trustees are personal information management systems, which are run by a neutral authority and are open to scientists to conduct research. There must be a more widespread discussion on data trustees, greater incentive for their development, and a different approach in the design of their responsibilities, according to Ms Specht-Riemenschneider, who concluded by emphasizing the importance of data access for research purposes as a key instrument for science and better policy development.

Mr Bolander of the Danish Technical University stressed furthering the capacity of AI to encompass social intelligence. According to Mr Bolander, **the best means to solve the AI dilemma is by developing algorithms with higher cognitive skills and that the real challenge with AI is not simply acquiring more data but developing more sophisticated and human-centric algorithms.** Mr Bolander explained that we will not develop fair and more moral AI by simply amassing more data. Instead, a greater effort should be made to develop systems with social intelligence. To achieve this, we need to have a better understanding of social intelligence itself, which calls for more interdisciplinary research in the development of AI algorithms. Mr Bolander concluded by stating that unless the AI systems are equipped with higher cognitive ability and more social functioning capacities, humans will likely remain wary of AI applications.

Ms Chander of European Digital Rights, an NGO, warned against the unchecked expansion of AI systems, stating that **promoting the uptake of AI should not be a policy goal in itself. According to Ms Chander, a blanket approach simply calling for more AI and more data may overlook the more complex issues linked to implementation, and an expansionist model of AI application could reinforce the dominance of big tech companies.**

Contributions by Political Groups

Disclaimer:
The statements herein below are drafted by the political groups of the European Parliament. The opinions expressed are the sole responsibility of the author(s) and do not necessarily represent the official position of the European Parliament. This document may contain links to websites that are created and maintained by other organisations. The AIDA Committee and the AIDA Secretariat do not endorse, nor are they affiliated with, the view(s) expressed in the said statements nor through the said websites.

The European People's Party Group (EPP)

AI is only as good as the data that it uses. The quality of this data, in turn, depends on good insights, and those are made mainly by data-literate and up-to-date experts. This is a challenge that needs to be tackled. What is also needed is a secure data sharing infrastructure and data sharing culture, and a good translation and transition from raw data to insights and from insights to action. The EPP is committed to a systematic and pragmatic approach to unleashing the potential of data that will nurture the development of AI. To this end, a strong digital alliance pact with the USA, based on a mutual understanding of AI, common ground on data protection and free data flows across the Atlantic, is also of key importance. To ensure the success of the EU data economy, education should also be a priority, as overcoming the digital skills gap, both in terms of acquisition of basic digital skills by the population and availability of highly skilled digital professionals, is essential for citizens and businesses to be able to embrace fully the opportunities of the digital economy.

The Progressive Alliance of Socialists and Democrats Group (S&D)

To boost a European AI according to our values and principles, the European data strategy must be human-centric, fair and socially sustainable, guided by the ethical principles of transparency, accountability, traceability, social responsibility, good governance, inclusion, privacy and human oversight. Europe needs a solid legislative framework to ensure that benefits are widely shared within the society and support safe, fair, legal and ethical sharing of data. Businesses should be encouraged to make data available to others, including through use of open source to promote interoperability, public scrutiny and citizens' trust that, in turn, will foster better training and validation of AI systems utilizing quality data.

We must ensure an evaluation of the fairness and quality of training data through the entire life cycle of AI: from data collection and selection processes, data security and protection measures, to data outputs. Citizens must be assured that their data will be collected and used in full respect of their fundamental rights. This is especially relevant in the context of data flows to third countries. For example, the EU-US Privacy Shield must uphold the highest standards in line with the European General Data Protection Regulation and the fundamental rights.

The Renew Europe Group

Access to data will be a cornerstone for strengthening the European industry through emerging technologies such as AI or the Internet of Things. We need to achieve a genuine single market for data and the EU data strategy will be crucial in this regard. The development of EU based cloud and computing capacities through initiatives such as the EU alliance for Industrial Data, Edge and Cloud or cross-border private initiatives such as Gaia-X are leading the way to build the foundational structures. Renew believes that the EU should maintain this fostering role.

Future data legislation will also be crucial to enable the seamless flow of data that AI requires. The Data Act should clearly define user rights, access rights and property rights in a fair and efficient manner taking account of co-generated data and IP rights.

We also need to break down data silos and enable cross-sectoral synergies if we want to tap the potential of data, still unexplored in most sectors. We need to facilitate cross-sectoral access to data and data portability and set our interoperability standards to compete on the international stage.

Greens/European Free Alliance

At the basis of the data economy lies the digitalization journey of every single business, start-up, SME, or large company. This requires availability of talent and research, a shift in business leaders' mindsets, the necessary digital infrastructure, and an enabling regulatory framework.

The Commission and Member States should ensure that existing market concentration in different layers of the data economy does not perpetuate itself in the European AI markets. Data access rights and portability rules should be designed to allow new market actors to access the data they need - and the market places where it is available - on fair, non-discriminatory and transparent terms, including with regards to pricing and ToS.

Clarification of access and monetisation rights to data, especially when co-created by multiple actors, as well as the enabling of multi-cloud strategies, are important components of a competitive data and AI economy. Moreover, the enforceability of metadata requirements to allow easy porting of data between different software applications, are needed to avoid a situation where the AI economy benefits only a few - already powerful - tech platforms.

ABOUT THE EDITOR:
Secretariat of the Special Committee on Artificial Intelligence in a Digital Age
Directorate General for Internal Policies of the Union

BRU - KOHL Building
aida-secretariat@ep.europa.eu

FOLLOW US:

www.europarl.europa.eu

@EP_ArtifIntel