DATA-DRIVEN ARTIFICIAL INTELLIGENCE FOR EUROPEAN ECONOMIC COMPETITIVENESS AND SOCIETAL PROGRESS

BDVA Position Statement
November 2018
INTRODUCTION

Artificial Intelligence (AI) has a tremendous potential to benefit European citizens, economy and society, and already demonstrated its potential to generate value in various applications and domains. From an industrial point of view, AI means algorithm-based and data-driven computer systems that enable machines and people with digital capabilities such as perception, reasoning, learning and even autonomous decision making. AI is based on a portfolio of technologies including algorithms for the perception and interpretation of vast amounts of information (data), software that draws conclusions, learns, adapts or adjusts parameters accordingly and methods supporting human-based decision making or automated actions.

One important driver for the emerging AI business opportunities is the significant growth of data volume and the rates at which it is generated. By 2020, there will be more than 16 zettabytes of useful data (16 trillion GB), reflecting a growth of 236% per year from 2013 to 2020. The Internet of Things (IoT) is driving this data explosion at unprecedented scales. Thus, IoT applications will need to analyse the vast quantity of Big Data and with recent advances in computing power and connectivity, more and more data can now be examined. AI is making great strides. In fact, according to IDC’s FutureScape: Worldwide IT Industry 2017 Predictions, by 2020, 40% of all digital transformation initiatives, and 100% of all effective data-driven IoT efforts will be supported by cognitive/AI capabilities.

This position statement expresses the view of the Big Data Value Association (BDVA) on Artificial Intelligence and Big Data. BDVA is an industrially led association with the objective to ensure Europe’s leading role in the data-driven world by fostering investments on technical and non-technical priorities along the data value chain. Given that data-driven approaches such as deep learning drove the recent breakthrough in AI, the BDVA is considered a strategic “Data for AI” partner in AI EU partnerships. This BDVA position statement primarily targets European decision-makers in the European Councils, the European Commission, the EU Parliament and European national government authorities. These are involved in shaping and planning of AI-related policies, European research programmes and funding instruments for Research & Development & Innovation (R&D&I).

AI OPPORTUNITIES

The current data explosion combined with recent advances in computing power and connectivity allows for an increasing amount of Big Data to be analysed anytime, anywhere. These technical advances enable addressing industrial relevant challenges and foster developing intelligent industrial application in a shorter time and with higher performance. AI will increase value creation from Big Data and its use to rapidly emerging B2B, B2G, G2C, G2B and B2C scenarios in many AI application domains. Machines and industrial processes which are supported by AI are augmenting human capacities in decision-making and providing digital assistance in highly-complex and critical processes.

Established industrial players are starting to implement AI in a wide range of industrial applications, such as complex image recognition, primarily for interpreting Computed Tomography (CT) and Magnetic Resonance Imaging (MRI); autonomously learning, self-optimizing industrial systems such as those used in gas turbines and wind farms; accurate forecasts of copper prices and expected power grid capacity.

---

2 IDC FutureScape: Worldwide IT Industry 2017 Predictions
utilization; physical, autonomous systems for use in collaborative, adaptive, flexible manufacturing as part of Industry 4.0 and many more. At their heart, many of these AI systems are powered by using data-driven AI approaches such as deep learning. Exploiting data ecosystems is essential for AI.

In addition to the above, the EU Big Data Value Public-Private-Partnership (BDV PPP) has established 32 projects with their respective experimentation playgrounds for the adoption of Big Data and AI solutions. In particular, the BDV PPP lighthouse projects play a fundamental role in piloting and showcasing value creation by Big Data with new data-driven AI applications in relevant sectors of great economic and societal value for Europe. These projects demonstrate the essential role of Data for AI:

**Data Bio:** *Data-driven Bioeconomy* takes on a major global challenge of how to ensure that raw materials for food, energy and biomaterials are sufficient in the era of climate change and population growth. Through Big Data and AI, DataBio is significantly enhancing raw material production in agriculture, forestry and fishery in a sustainable way. With its 26 pilots, DataBio strives to demonstrate annual increases in productivity ranging from 0.4 % in forestry to 3.7 % in agriculture and fishery (through savings in vessel costs). This makes up for a productivity gain of 20 % over five years in agriculture and fishery. Big Data pipelines and AI techniques are used in multiple pilots using the DataBio platform deployed in multiple clouds. The platform gathers Earth Observation data from satellites and drones as well as IoT sources from in-situ sensors in fields and vehicles. It manages and analyses the generated Big Data and presents it to the end-users. These include farmers, foresters, fishers and many other stakeholders in a user-friendly way for supporting their operational decision making by providing guidance in critical daily questions, such as what and where to grow, crop or fish; how to fight diseases; or when and how to harvest, cut or fish.

**TransformingTransport** demonstrates in a realistic, measurable, and replicable way the transformation that data-driven AI solutions can bring to the mobility and logistics market in Europe. Mobility and logistics are one of the most used industries in the world – contributing to approximately 15% of GDP and employment of over 11 million people in the EU-28 zone, i.e. 5% of the total workforce. With freight transport activities are projected to increase, since 2005, 40% in 2030 and 80% in 2050. This will transform the current mobility and logistics processes to significantly higher efficiency and more profound impact. Structured into 13 different pilots, which cover areas of significant importance for the mobility and logistics sector in Europe, TransformingTransport validates the technical and economic viability of big data-driven solutions for reshaping transport processes and services across Europe. To this end, TransformingTransport exploits access to industrial datasets from over 160 data sources, totalling over 164 TB. Initial evidence from TransformingTransport shows that Big Data-driven solutions using AI may deliver 13% improvement of operational efficiency. The data-driven solutions in this project entail both traditional AI technology for descriptive analytics (such as support vector machines), as well as deep learning methods employed for predictive analytics (such as recurrent neural networks). With today’s promising results using AI technology (e.g. 40% increase of prediction accuracy), we expect such AI solutions of advanced analytics as an enablement to automated decision support for operational systems. These will establish the next level of efficiency and operational improvements in the mobility and transport sectors in Europe.

**BigMedilytics:** In 2014 the EU-28’s total healthcare expenditure was €1.39 trillion. Spending is expected to increase to 30% by 2060 primarily due to a rapidly ageing population who typically suffer from chronic diseases. These figures indicate that current trends within the EU’s Healthcare sector are very unsustainable. The BigMedilytics Healthcare Lighthouse project demonstrates how the application of AI technologies on Big Data can help disrupt the Healthcare sector so that quality, cost, and access to care can be improved. Market reports predict a CARG of 40-50% for AI in Healthcare, with a market size to euro 22B by 2022. The project applies data-driven AI technologies over 12 pilots which focus on three main themes: 1- Population Health; 2- Oncology; and 3- Industrialization of Healthcare. These effectively cover

---


4 According to the ALICE ETP, a 10% efficiency improvement will lead to EU cost savings of 100 B€
AI OVERARCHING CHALLENGES

1. Business models

With the recent technical advances in digitalisation and AI, the real and the virtual worlds are continuously merging, which, again, leads to entire value-added chains being digitalised and integrated. For instance, in the manufacturing domain, all the way from the product design through to on-site customer services is digitalised. The increase in industrial data combined with AI technologies triggers a wide range of new technical applications with new forms of value propositions that shift the logic of how business is done. To capture these new types of value, data-driven AI-based solutions for the industry will require new business

5 Big Data PPP selected project from the ICT-11.a call 2018
models. The design of data-driven AI-based business models needs to incorporate various perspectives ranging from customer and user needs and their willingness to pay for new AI-based solutions to the data access and the optimal use of technologies while taking into account the currently established relationships with customers and partners. Successful AI-based business models are often based on strategic partnerships with two or more players based on transparent ways of resource-, investment-, risk-, data- and value-sharing establishing the basis for sustainable win-win situations.

2. Trust in AI

With AI’s disruptive potential, there are significant ethical implications on the use of AI & autonomous machines, and their applications for decision-support. Future AI research needs to be guided by new established ethical norms. Although the current AI methods have already achieved encouraging results and technical breakthroughs, results in individual cases show some concerning signs of unpredictable behaviour. Recent studies showed that state-of-the-art deep neural networks are vulnerable to adversarial examples or are unable to cope with new unknown situations. To overcome those shortcomings, for any critical applications (where “critical” needs to be defined with clarity), one should be able to explain how AI applications came to a specific result (“Explainable AI”). Explainability will ensure the commitment of industrial users to measurable ethical values and principles when using AI. One should foster responsible technological development (e.g. avoid bias) and enhance transparency in such exercise. Explainable AI should provide transparency about input data as well as the “rationale” behind the algorithm usage leading to the specific output. The algorithm itself need not necessarily be revealed in this case.

The purpose of AI, data analytics, machine and deep learning algorithms is not only to boost the effectiveness and quality of the services which are delivered to the client, but also ensure that no negative impact is brought as a result of deploying AI solutions in critical applications. For instance, ensuring that AI-powered systems treat different social groups fairly is a matter of growing concern for societies. FATML, i.e. Fair Accountable and Transparent Machine Learning, is an emerging important multi-disciplinary field of research. Related areas including Big Data for social good, humanistic AI and the broader field of AI Ethics have only recently started exploring complex multi-faceted problems, e.g. fostering the creation of social and human-centred values by adding new parameters and enhanced objective functions and restrictions.

Trusted AI involves the simultaneous achievement of objectives that are often in conflict. One critical challenge stems from the ever-increasing collection and analysis of personal data and the crucial requirement for protecting the privacy of all involved data subjects as well as protecting (commercially) sensitive data of associated organisations and enterprises. There are some approaches attempting to address this issue, including security-oriented (e.g. machine-learning on encrypted data with secure computation technologies), privacy-enhancing (e.g. detect privacy risks and alert users) and distributed processing ones (e.g. federated machine learning). As all privacy approaches add cost and complexity to AI systems, the optimal trade-offs without adding considerable complexity are essential research challenges to be addressed. A critical problem is presented by the difficulty to allocate and distribute liabilities and responsibilities across assemblages of continuously evolving autonomous systems with different goals and requirements. While existing risk-based, performance-driven, progressive, and proportionate regulatory approaches have promised a more flexible, adaptive regulatory environment, stakeholders are increasingly struggling to deal with the complexities of multi-level, multi-stakeholder and multi-jurisdictional environments within which AI is being developed. Multidisciplinary efforts at both international and regional level are therefore required to ensure the establishment of an enabling

environment where trust and safety of AI are dealt with from a global governance perspective. Existing tools from other domains, such as regulatory sandboxing in fintech, testing environments for autonomous vehicles and so forth, could serve as incubators for establishing new policy, legal, ethical and regulatory norms and measures of trusted AI in Europe.

3. Ecosystem

For developing sustainable data-driven AI businesses, it will be central to consider a value-network perspective, i.e. looking to the entire ecosystem of companies involved in value-networks. The ecosystems will be increasingly shaped by platform-providers who offer their open platform based on open standards to their customers. European economic success and sustainability in AI will be driven by ecosystems which need to have a critical size. Speed is necessary for the development of these ecosystems.

Data sharing and trading are essential ecosystem enablers in the data economy, although closed and personal data present particular challenges for the free flow of data. The EU has made considerable efforts in the direction of defining and building data sharing platforms. However, there is still a significant way to go to guarantee AI practitioners the access to large volumes of data necessary for them to compete. Further actions must be carried out to develop Data for AI platforms, such as awareness campaigns to foster the idea of sharing their data in companies and research centres, and incentives for parties to join data exchange/sharing initiatives. To overcome barriers to data sharing for AI, frameworks for data governance need to be established to enable all parties to retain digital sovereignty over their data assets. Obviously, sharing data must be done, from the legal point of view, preserving privacy by anonymizing all the attributes referring to people, and respecting commercial interests (IPR, competition, ownership) by providing solutions to deal with technical and legal challenges such as data governance and trust-enhancing protocols for data sharing/exchange, decentralized storage and federated Machine Learning. And from the technical perspective, by (1) designing information systems (i.e. databases) in order to ensure the future use of the datasets with minimal efforts in terms of cleaning data or defining ontologies, by (2) transforming/mapping data sources taking into account the variety and heterogeneity of data in order to gain interoperability, and (3) by ensuring the veracity of shared data according to quality standards.

Open AI platforms will play a central role in the data economy at three different levels: (1) definition of protocols and procedures for uploading datasets into data sharing platforms, (2) definition of standard APIs for different libraries (AI/ML, image processing, etc.), and (3) the design and development of a web-based user interface to allow data scientists to upload data, define pipelines of transformations to apply to data before training and testing AI models, and to choose among a wide range of AI techniques to run on the same data in order to carry out comparative studies. Successful European Open AI platforms require the contribution of many agents, such as universities, research centres, large companies, and SMEs.

By relying on data-sharing platforms, data innovation spaces, open AI platforms and digital innovation hubs, industrial collaborations between large and small players can be supported at different levels: technical, business model and ecosystem, while, at the same time, ensuring data and technology access for SMEs and Start-Ups. To complement technical and legal infrastructures for the free and controlled flow of industrial data, the building and nurturing of industrial ecosystems fostering data-driven industrial cooperation across value chains and therefore networks will have a critical impact.

Enabling data-driven AI-based business models across value chains and beyond organisational boundaries will significantly maximise the impact of the Data Economy to power European AI industries. Mechanisms that overcome the lack of data interoperability and foster data sharing and exchange need to be defined

and implemented. Notwithstanding, the creation of and compliance with binding international standards is of central importance to the sustainability of solutions and thus it is a competitive strength. Preferably these standards should be global – because only global standards ultimately lead to success in a world that is more and more networked and where multinational companies make significant contributions to national GDPs.

4. Technology

Success in industrial AI application relies on the combination of a wide range of technologies, such as:

**Advanced Data Analytics:** Many data analytics techniques require adaptation for running more efficiently when working with large datasets. These improvements rely on the development of new algorithms and new ways of transforming data. Additionally, with self-adjusting AI systems, machines will become self-operating by making decisions according to specific contexts to dynamically optimise performance, beyond the level of efficiency the same AI systems can reach when adjusted by humans.

**Hybrid AI:** To derive value from domain knowledge, methods from both symbolic AI and statistical AI need to be combined to give the maximum potential and usability of AI-based applications. This combination for making use of knowledge graphs and statistical AI techniques supports AI solutions with regard to (i) data quality issues, (ii) better integration and use of training data, (iii) explainable AI (no black-box solutions) and, finally, (iv) the mutual fertilisation of semantic technologies and AI techniques towards self-optimising machines.

**Distributed AI / Edge Analytics:** The increasing number of intelligent devices at the edge is one of the critical elements for AI. We are now at the point where the collective computing power at the edge (outside data centres) is surpassing the centralised computing capacity. As computing capabilities in the cloud and at the edge are increasingly intertwined, we will see the emergence of distributed AI, i.e. new research approaches that will bring the AI to be the core of most future data analytics-based applications.

**Hardware optimised to AI:** Specialized hardware devices and architectures have an increasingly strong impact on both the AI learning process on applications with large datasets, and on the predicting/inference task. In particular, when fast decisions and actuation matters. The designs of powerful and affordable systems in both sides of the AI dataflow are an important research topic. And in addition, the other way around, AI algorithms need to be optimized to take advantage of hardware accelerators.

**Multi-lingual AI:** Humans use language to express, store, learn and exchange information. AI-based multilingual technologies can extract knowledge out of tremendous amounts of written and spoken language data. Processing of multilingual data empowers a new generation of AI-based applications such as question answering systems, high-quality neural machine translation, speech processing in real time, contextually and emotionally aware virtual assistants for human-computer interaction.

BDVA RECOMMENDATIONS

1. **Set-up a European Public-Private-Partnership on AI (Partnership)**

Leverage existing industrial and research communities and build upon results of the Big Data Value PPP and the SPARC PPP to establish a European Public-Private Partnership on AI (a PPP on AI) in collaboration with the European Commission and the Member States. Sturdy and stable long-term collaboration structures need to be put in place to develop the European AI Ecosystem that will accelerate value creation by AI in European Businesses, Governments and Society.
A PPP is needed to realise the economic value of AI in private and public sectors (and in particular in the B2B world) fostering industrial competitiveness and modernization, and, societal transformation. Digitalisation of private and public sectors offers a unique set of opportunities for Europe to remain a global leader. European companies have the opportunity to strategically exploit Big Data powered AI technologies to drive current and future business values, for the ultimate benefit of European society as a whole.

2. Develop the European AI Ecosystem (Ecosystem)

Active engagement of all relevant stakeholders in designing policies and supporting programmes to promote the adoption of AI technologies in all industrial sectors. Of utmost importance, furthermore, Europe needs a holistic AI strategy that encompasses the necessary framework for implementing legal, regulatory, communication, R&D&I and other relevant actions.

Scale Industrial Cooperation Models in the Data Economy by building on data-sharing platforms, data innovation spaces, and digital innovation hubs, industrial collaborations between large and small players (at the technical, business model and ecosystem levels) can be supported, while at the same time ensuring data and technology access for SMEs and Start-Ups.

Help small and medium businesses to understand and use the potential of AI. Specific action is needed for bringing SMEs on board in industrial domains, since the importance of digitalisation is still not widely known/accepted, especially by non-ICT SMEs which need AI-based solutions in their business processes.

Promote AI using positive stories/use cases.

3. Boost the European AI Ecosystem (Ecosystem enablers)

Data for AI: To unleash the power of AI, we need to further enable and facilitate access to data for European industry, academia and research institutes, taking into account the principles of transparency, shared value creation and data governance frameworks that respect all parties’ commercial interests, ensure undistorted competition and minimized data lock as well as B2G data sharing principles. This will involve policies, regulations, business models and technology innovations in the fields of data platforms and enabling data technologies such as privacy-preserving computations.

Investment on Research, Development and Innovation for AI:

- **Focus R&D&I investment and policy support on AI to industrial domains** to strengthen and ensure Europe’s leading role, e.g. combining automation, AI, semantics, edge computing (data) analytics and cybersecurity, to take the efficiency of industrial infrastructures (factories, power, transportation, etc.) to the next level. It is of utmost importance to align and coordinate EU and member states AI initiatives to collaborate tightly with a concerted support effort from the EU and industry leveraging the complementary strengths. Large-scale AI research and industry and science-driven innovation digital hubs (CERN-like) should be established in Europe that can compete with the technological advancements in the United States and China. With the creation of industry and science-driven digital hubs, European AI talents and relevant stakeholders will be able to bundle forces for fast innovation and avoid dispersed efforts.

- Research is required to increase the trust in (i) algorithms and data, in (ii) the co-evolution between humans and AI-base systems, and in (iii) the legal and ethical challenges associated with making data-driven critical decisions. This research for increasing the overall confidence on AI-based solutions is crucial for reaching the global governance of AI.
Next-Generation Data and Artificial Intelligence Platforms (technology): Europe needs to ensure it obtains a leadership position in the development of next-generation Data for AI platforms and the advanced data management techniques necessary to foster the continued growth of European Data Economy as an enabler of AI-based European Industry. AI needs to evolve in parallel with the technologies adopted in cybersecurity, to achieve an environment of trust in a digitally connected, AI-driven world. AI evolution should be by design intertwined with cybersecurity.

Skills: Europe needs to ensure it retains the best educators and academic delivering educational offerings to the next generation of data scientists, data engineers, and data workers. However, talented computer scientists are being lured from academia by the private sector with negative consequences for future teaching and research. Europe needs to work to reverse the AI brain drain from academia by supporting academic/industrial mobility and driving new forms of academic and industrial research and educational partnerships. Broad, cross-domain understanding will be needed in addition to deep single domain expertise. Besides education, vocational training also must be taken into account.

Flexible Legal, Regulatory and Governance Frameworks: Building trust is dependent upon a firm, yet flexible regulatory and governance framework which delivers legal certainty and predictability for stakeholders as the ultimate goals of law and regulation. While AI is a transversal technology, it is nonetheless applied differently in different application scenarios. Thus, a horizontal principle-based framework should be complemented by legal and regulatory amendments in various (safety-critical) vertical domains, such as transportation, healthcare, etc., which are subject to multilevel and often multijurisdictional requirements determined by the specifics of the respective domain(s). It is essential that the evolution of the legal, regulatory and governance frameworks should not occur in isolation from the development process. It is therefore essential to establish a favourable environment where novel technological and regulatory solutions could be tested, improved and verified, ensuring thereby the sustainability of the existing governance mechanisms and the values of legal certainty and predictability.

ABOUT BDVA

The Big Data Value Association (BDVA) is an industry-driven international not-for-profit organisation with 200 members all over Europe and a well-balanced composition of large, small, and medium-sized industries as well as research and user organizations. BDVA is the private counterpart to the EU Commission to implement the Big Data Value PPP program. BDVA and the Big Data Value PPP pursue a common shared vision of positioning Europe as the world leader in the creation of Big Data Value.

The mission of the BDVA is to develop the Innovation Ecosystem that will enable the data-driven digital transformation in Europe delivering maximum economic and societal benefit, and, achieving and sustaining Europe’s leadership on Big Data Value creation and Artificial Intelligence.

BDVA enables existing regional multi-partner cooperation, to collaborate at European level through the provision of tools and know-how to support the co-creation, development and experimentation of pan-European data-driven applications and services, and know-how exchange.

For further information: www.bdva.eu / info@core.bdva.eu /@BDVA_PPP

Editors: Sonja Zillner (Siemens), Jon Ander Gómez (Universitat Politècnica de València), Ana García Robles (BDVA), Edward Curry (Insight).

Contributors: Nozha Boujema (Inria), Andreas Metzger (paluno, Univ. Duisburg-Essen), Zoheir Sabeur (University of Southampton), Martin Kaltenböck (Semantic Web Company), Thomas Hahn (Siemens AG), Marija Despenic (Philips), Caj Södergård (VTT), Natalie Bertels / Ivo Emanuilov (imec-CITIP-KULeuven), Simon Scerri (Fraunhofer), Andrejs Vasiljevs/Tatjana Gornosttaja (Tilde), Axel Ngongo (Technical University of Paderborn), Freek Bomhof (TNO), Yiannis Kompatasariadis / Symeon Papadopoulos (ITI Greece), Juan-Carlos Perez-Cortes (ITI Valencia), Oscar Lázaro (Innovalia Association), Aitor Arnáiz (IK4-Tekniker).