AI Deep Dive Section

1: Overview of Artificial Intelligence

1.1 Artificial Intelligence

At the core of AI as a field is developing techniques to perform tasks that we normally associate with the use of intelligence. That is the reason why mimicking tasks such as learning, reasoning, planning and interacting with other intelligent agents are of interest – as well as addressing challenging tasks like understanding human language, intentions and actions, along with solving hard problems such as scheduling, theorem proving and constraint optimisation. In basic science, this leads to a quest for an understanding of the mechanisms underlying thought and intelligent behavior and their embodiment in machines. In applied science, this leads to a quest for new methods for solving difficult problems. In business this leads to a quest for combining these advancements with increasingly better, cheaper and more available hardware, software, infrastructure and data, into products, services and new market opportunities.

All machine learning or AI-systems today pertain to weak AI. They can perform single and rather narrowly defined tasks, and we are far today from artificial general intelligence. Some first steps are being proposed, addressing the combination of learning and reasoning, multi-task and multi-modality learning, as well as life-long learning. A long way is still ahead, scientifically speaking, for instance to reach common sense reasoning, but we should nevertheless already consider the moral and ethical implications of research aimed at AI systems that display the full spectrum of human intelligence (including consciousness).

Deep learning is the capability to input raw data as they are observed with the desired outcomes to learn how to predict that outcome for unseen cases. Machine learning has recently revolutionized various fields of science and has also started to pervade commercial applications in an unprecedented manner. Machine learning has evolved into well-founded principles and software suites to support it, and learned from data, where most of the impact is still ahead of us.

At the same time, automated reasoning forms the basis for the correct and trustworthy operation of all modern computer hardware, and increasingly, also of software. AI optimisation techniques are widely used to ensure efficient use of resources in business, industry and public administration. These developments are certain to increase in impact as all of these areas of AI are developed further.

1.2 AI in context

While the field of artificial intelligence (AI) has a rich history of more than 60 years, it is the last decade or so where major success has been achieved. While AI's various subfields have seen essential developments, also external factors have played a role. This includes the emergence of better hardware, the internet, mobile phone networks, cloud computing, and containers. It also includes an impressive growth in the production of digital data and access to them; as well as a substantial democratization of technology driven by falling prices, miniaturization, advances in human-computer interaction, and available software tools and components. The examples also include the significant ease of deployment as society and its infrastructure have become increasingly digitised. Another key to the recent success of AI has been on the demand side. As society's grand challenges are becoming increasingly pressing, from resource limitations, climate change, pandemics, and unstable global financial instruments, to autonomous weapons, security and migration pressures, scientists are asked to help.

While these developments give an explanation for the increased return on AI investment, and the scientific and commercial success of AI the last ten years, they do not explain the extent of the interest in the field. The explanation for that is rather that these developments finally have brought the game-changing promises of AI within reach. Many argue that we are entering a new industrial *revolution*, or in an attempt to be more precise, a technology-driven societal transformation.

In this transformation, *industrial sovereignty* is at stake. Recent geo-political developments have demonstrated once more that industrial sovereignty must be a cornerstone of Europe. It is upon sovereignty that a European ethical approach, that is human-centric and trustworthy, can be built. For AI, sovereignty is not a simple thing to achieve, as both digital data and digital products are boundlessly fleeting over the world. AI-services can be run from everywhere to everyone. These services are accumulating data we have handed over, for others to use AI-technologies to draw strategic consequences. At one point in time this digital-data-trade-deficit may come back to Europe. Then, close ties of fundamental AI research and development to EU-based manufacturing and EU-based service industries to reduce the need to separate data from the product.

Technology has always defined the reality we live in. AI turns this definition, this impact on the way we live, work and think, even deeper.

While major nations across the world are significantly investing in AI, *Europe is facing major barriers* in embracing it. First and foremost, AI is driven by talent. However, Europe seems to have difficulty retaining top talent in European institutions and industries alike. Salaries are not competitive, risk capital is sparser, and scale-up competence even sparser. There is a less fluid relationship with industry. AI technologies have considerable potential to disrupt existing markets. Increasing the entrepreneurial capacity and volume is therefore a crucial element in the AI landscape. However, there is considerably less private capital and competence available for starting and scaling up new enterprises. And there is a lack of critical mass in research and connection to fundamental research such that technology-driven economic and societal impact lags behind other continents. If Europe will not embrace AI, it will come from the other two continents as AI is boundless.

The abundance of *digital data* in combination with learning have redefined the relationship of industry to society and vice versa. In new territory of privacy law, Europe has defined GDPR as a first attempt to protect personal digital data. In a new departure for tax law, several states in Europe are considering taxing digital data. The European parliament has accepted a motion to curb personalised advertising on the basis of data derived from cookies and the like. It is evident the landscape is changing fast, and that these elements cannot be set apart when considering AI-technologies.

The objective of many AI-system is the taking of an action. Actions can be physical as in *robotics*, where machines construct a product, or they fulfill a task in logistics. And actions can be non-physical as in search robots, gathering information from across the Internet. As for AI, the underlying principles, whether the action is physical or non-physical are the same: AI-algorithms reason towards goals.

Economically, a key competitive factor in the success of the EU is how quickly Europe will be able to use data broadly and efficiently. Trustworthiness is a crucial aspect of that. Learnability is another aspect, already essential in some B2B-applications in Europe. For some applications, human-centric machine learning and decision-making, including fairness and explainability, are important. For all of this, talent is the main driving force – for industry, academia and public administration. Even for societal impact, talent is key.

AI has emerged with the power *to innovate the systems of industry and society*. It is hard to decide which *new application areas* will show up over the next ten years. Generally, we will see that the potential range of AI-based systems and products is as broad as the range of human activities. We find systems or products in areas as different as cybersecurity transportation, emission reduction, efficient use of resources, autopiloting, robotics, gaming, finance, transport, marketing, industrial production and public administration. It follows that AI is not a technology or field of science, but rather a broad range of research and technologies with an unusually wide range of applications.

Europe’s future, its ability to tackle environmental and societal challenges while also strengthening European competitiveness and positioning the continent in the global knowledge economy, is dependent on its ability to establish well-functioning ecosystems of research and innovation. Successful ecosystems have strong collaboration between different sectors, including businesses and industry, entrepreneurs and startups, investors, academia, and governments.

Many *start-ups in AI* originate from academia, some of them will grow fast. A fascinating new phenomenon is that in AI and in machine learning the borders between academia, industry and services are transforming. Europe trails behind in this respect in what other continents are achieving. To stem the data and brain drain, Europe needs to put talent first, start-ups and scale-ups need to be fostered, services and industries need to be helped to adopt and make efficient use of the new technology, and societal concerns need to be dealt with at the core of the technology. All of this needed to provide economic stimulation, and to ensure European technological sovereignty.

1.3 Market perspectives of AI

In a recent communication by the European Commission [1], a Europe-wide survey was conducted on *the uptake of artificial artificial intelligence* among enterprises. Early 2020, the survey reached a total of 9640 enterprises. It measured a high awareness of AI for three quarters of all enterprises, where half have adopted at least one AI-technology and a quarter have adopted two or more. In fact AI-adoption was found to be highest in the ICT-sector and finance while it was low in construction. While one in five have plans to adopt AI in the next two years, two in five have neither adopted AI nor plan to do so. Adoption at the level of each technology is still relatively low: from 3% for sentiment analysis to 13% for anomaly detection and process or equipment optimisation. The most common sourcing strategy is external, as 60% of EU-enterprises that use AI purchase software or ready-to-use systems. For half of all enterprises, key *internal* barriers to AI-adoption are difficulties in hiring new staff with the right skills, the cost of adoption, and the cost of adapting operational processes. One third of all enterprises find liability for potential damages, data standardisation, and regulatory obstacles, to be major, but less important, *external* challenges to AI-adoption.

The study demonstrates the importance of the AI-wave of innovation as it goes through three quarters of all enterprises in Europe. he study demonstrates that it is important to remove internal obstacles due to lack of talent, and external obstacles by providing certainty on regulation.

Industry and services now and in the near future rely on AI-technology, inevitably. An important driver of the current intense innovation wave is the pervasiveness of digital data. And, as a consequence of the abundancy, research in AI has emerged as a motor to drive innovation in industry and in services. The quality of that connection will determine the success of a continent in the innovation wave which is currently underway. The more industries and services rely on data, the more they need to embrace the opportunity for business model innovation as well.

Banking has been disrupted by new forms of money, new forms of payment and new forms of risk assessment, new forms of fraud. This requires cyber security, which increasingly is based on AI techniques but also AI-techniques like machine learning, which are essential to predict the solvency of clients while ensuring proper fairness. New billion Euro companies have grown with new business models such as specialized but grown-big fin-tech companies, and also big-tech have ambitions to shake classical models.

Retail has been disrupted by new forms of home delivery, new forms of advertising and new forms of routing. The business models of the sector are transforming themselves necessarily for survival due to the entrance of young giants as Amazon. New billion Euro companies have grown and old retail is transforming itself speedily. The success in this innovation will be determined by the effectiveness of AI-techniques like machine learned personalisation under proper protection of privacy and detailed planning stock and transportation.

Restaurants and personal services are being disrupted by new forms of delivery by new billion Euro companies such as Just Eat Takeaway and Deliveroo, also in response to Uber. The business models have not found an equilibrium yet, but the success in this innovation will be determined in part by the effectiveness of new business models integrating AI-techniques like machine-learned and machine-optimized personalisation under proper privacy protection and detailed prediction of societal moods. From many new internet companies we have seen how a sector will be transformed by re-inventing the digital information flow coupled to predicting where and when it will be needed by AI-techniques. Apart from information standards and chain normalisation this requires AI to survive.

In *medicine and care*, the integration of AI into the daily work is already on its way. This will require AI-techniques for AI-assisted diagnosis, AI-based monitoring, abnormality signaling and early warning for proper care. This will require AI technologies like deep learning, verification, and optimisation, and a deep and proper attention to personal protection deeply integrated into the technology. European companies in healthcare are transforming into AI-companies gaining experience from digital dossiers and digital medical knowledge to personalize medical care. More than in other application areas, explainability of decisions is of importance here, as well as provable safety and correctness of AI-driven systems. They will be supplemented with new companies for pharmaceutics discovery through reinforcement learning and other AI techniques, as for instance healthcare monitoring under anonymity.

Many *service sectors* rely on handling information and knowledge effectively and efficiently: notary, law, real-estate, transport, trade. This has caused the emergence of a new sector: information search, trend prediction and strategic information management all based on data-driven and model-driven AI techniques together with effective techniques for optimisation and natural language processing. Where data-tech companies dominate the market, Europe cannot afford to lock itself out of this sector. We see the emergence of specialised billion Euro companies such as for literature and for judicial information services.

The above list supplements the emergence of AI in data analysis, manufacturing, and agriculture where AI-branches of machine vision, human-machine interaction, planning, reasoning, quality control, and feedback are already coming into use.

1.4 Example of areas of work

The work of the PPP can be organized by grouping sectors of industry according to the similarity of their needs and define PPP AI-programs for them. By example, banking and the financial sector will employ AI and cyber security, large data learning, fairness in AI. Retail will need AI-planning and routing, AI and personalisation. Personal services will profit from limited personal data flows, personalization, information retrieval. Medicine and healthcare will be boosted by AI-assisted diagnosis, privacy, small data learning, explainable AI, care surveillance. Professional services will need NLP, AI-guaranteed search, explainability, trend prediction. The newly forming sector of digital data handling will need algorithms to come to data, precise data, knowledge engineering. Manufacture and production will use action, planning, sensor data, AI quality control, autonomy. Agriculture and the food chain will employ robotics, sensor data, AI Quality control, autonomy. In arts, entertainment and cultural production AI will be employed. In earth- and space science will need computer vision and trend prediction. Education and Training will turn digital and intelligent by personalisation learning and reasoning. Energy and resources will be used more efficiently by planning. The environment can be protected more effectively by modelling and planning. Information and Communication Technology will be more efficient by intelligent routing. There are many examples in public administration and citizen services and there are many opportunities being employed in science, innovation and design. Transportation will reach autonomous vehicles as only one example in that branch of the application of AI.

2. Framework and enabling factors

Commercial and scientific progress on AI is not only a function of scientific development, but on underlying infrastructure, regulatory environment, household digitalization, and governmental policies.

2.1 European framework

2.1.1 European rights, principles, and values in AI

Responsible research and innovation. Utilizing tangible results of the Horizon2020 program in responsible research and innovation helps guide Europe's competitive effort in the field of AI towards socio-economic benefits in Europe, democratic development, rule of law and human rights.

Explainable AI. It is generally accepted in both research and industry that explainable AI is an important element. In the public debate on AI it dominates as a way to enforce a reasonable and ethical version of AI. Explainable AI is necessary to engage the public in the debate of the values, changes and opportunities AI brings, and explainable AI is sometimes an essential element by itself. There are also large sectors of society where explainable AI does not play a role. When the performance of an AI-application is predictable or the application is under strict human supervision or the application has no societal harm, the most accurate result will prevail and rightfully so. In manufacturing, a key asset of Europe's economy, the best performing AI-algorithm will generally win. Enforcement of explainable AI should be differentiated to the use and goal of the application. Where explainable AI is needed, a tendency is observed to equate explainable AI with symbolic AI as if only explicit rules can deliver explainability to AI. There are other ways of explanation: by visualization, by visual examples as is the preferred way of teaching in medical radiology, by circumscription of abstract notions such as "justice" or "relevance". It is important to assess where explainable AI should be enforced and what are the permissible ways to achieve that.

Trustworthy AI. The European Commission, building on the work of the High-Level Expert Group on Artificial Intelligence, is focusing strongly on the development of Trustworthy AI. In their recent Whitepaper on AI they further built on this concept with a vision for regulation in AI. Trustworthy AI systems must be ethical, legal, and robust. A framework of seven key requirements have been developed to characterise the notion of trustworthiness more precisely.

2.1.2 Promoting value for business, society and people in AI

AI for good. In order to strengthen political support for AI in Europe, it is useful to focus the European strategy on how AI can be used for improving people's lives. AI for good is employing AI to tackle some of the world's greatest economic and social challenges, AI for All. Global leadership in this focus can make the term "AI made in Europe" synonymous with commercial products and services bringing AI for good. Again, the European focus here is on Trustworthy AI (as discussed above).

AI-innovation ecosystems. Strong AI-innovation ecosystems are a prime enabling factor. When looking around the world, global success in AI requires that all parts of the AI-ecosystem are well developed and interconnected: entrepreneurs, innovators, investors, regulators, large and small enterprises, and applied and foundational scientists and research teams. Innovation happens increasingly as non-linear interactions between the actors in the ecosystem (in contrast to purely linear technology transfer from university to industry), a development a shift that is strengthened by AI.

Academic collaboration with industry. It is critically important that industry-academic collaboration is encouraged and structured using transparent and simple IP rules that ensure that public funding is used in a way that benefits the public. Joint research involving industry and public funding should be published openly. Crucial in AI and especially is to stimulate an active start-up culture. AI-researchers should be encouraged to found startups. The emphasis should not be on the optimization of short-term licensing income but be aimed at sustained impact, thus generating downstream impact in Europe. We should facilitate startups in many ways.

2.1.3 Policy, regulation, certification and standards in AI

Regulation and market. Competitive use and uptake need both regulations and functioning market structures. Clearly regulations and policy instruments can both hinder and advance Europe's AI developments, including markets, new research, and applications.

Regulatory sandboxes. The European Commission, the European Parliament, and various national AI-strategies have recognised regulatory sandboxing as a strongly enabling factor technologies and, most prominently, AI.

Governmental funding instruments. With the weaker private investment in AI in Europe, governmental funding instruments become that much more important. The European Commission and the Member States are working through a framework envisaged in the Coordinated Plan for AI[1] that ensures alignment between Member State strategies in AI. The PPP will provide an important level of complementarity to those efforts by connecting and supporting the Commission's Digital Innovation Hubs, and opening them up to the wider European AI innovation ecosystems.

Privacy and safety by design. Research and development in the areas of privacy- and safety- by design in the last two decades provide a market opportunity for European businesses. Such a position can be strengthened by the more recent work in ethics by design for AI. The European view for digital privacy as currently regulated in GDPR is a good first step. A commonly felt weakness would be if regulation separated from the AI-technology, opening the door to circumvention so that there exists an easy work around of the regulation. Privacy and ethical considerations, in general, can best be studied in connection to developing the best techniques of AI. An opportunity would be to integrate privacy into the machine learning technology, for example by the application of a cryptographic technique, now popular in machine learning, known as differential privacy, aiming to guarantee an explicit and limited list of conclusions to be allowed from data, or by applying watch-dog technologies.

Data accumulation. Too little attention has been paid in society to data accumulation and the dangers it brings, especially when it flees across to other continents. Accumulation of data from all corners of life is an unrated weakness of our current digital world as it could easily jeopardize the anonymisation of data as framing would easily fill back in the name and other ID-information removed from the data. Another approach to put limits to accumulation is federated machine learning, applying algorithms to the data to learn without the need to first centralise sensitive data and still learn for a specific purpose. Therefore, Europe would want to endorse the ethical, explainable, data accumulation and privacy constraints in close combination with the further development of the AI-technology.

Regulation and certification of AI. Regulations by means of developing standards, metrics, legislation and institutional mechanisms for auditing, monitoring, inspection and certification, is an important context for Europe's AI strategy. Regulations can, when designed right, boost European competitiveness. In regulation, verifiable AI is important. It is a natural task for the EU to regulate. It is important to certify machine learning in safety critical systems on SIL-levels. Being able to certify the safety level offered by any machine learned or complex knowledge-based algorithm has great impact, possibly more impactful than explainability of its decisions. The alternatives of formal methods of verification are intrinsically difficult to generalize to data heavy applications. Considering that Europe's economy is significantly based on building machines, from automotive to industrial robotics and manufacturing in general, and considering that putting AI into these machines would likely have a strong impact on the EU economy. AI4EU could play a role here. An opportunity arises for Europe to stimulate good methods and procedures for verification, which for machine learning would surpass formal methods of verification and include methods, which can handle external (training) data.

2.2 Enabling factors

2.2.1 Skills and knowledge in AI

First and foremost, *talent* will enable the growth of AI, as was also found in the recent study on AI-adoption [1]. The spread of fundamental knowledge is essential for the continent, within academia and across to industry and back. Visiting researchers from academia and industry as well as workshops and summer schools for students, academics, and industrial participants help to promote mobility, facilitated by housing, childcare, and international schools at each site. AI made in Europe is equivalent to fostering talent to stay in Europe and

[1] https://ec.europa.eu/knowledge4policy/publication/coordinated-plan-artificial-intelligence-com2018-795-final_en

pursuing competitive AI-research at the world-level with best opportunities for societal and economic impact. Investment in AI talent is not a free choice. When first talent and then data of all sorts are shipped to other continents, one day this will have consequences for the AI-models that will make our economy, culture and society efficient. In order for Europe to have a say in the values important to her.

2.2.2 Data in AI

Data. The significant increase in data volume will continue. Europe's data production does not guarantee global competitive strength in itself. Strong enablers to AI include providing access to data and data feeds from the increasing number of data producing objects in society, regulate its use, and develop functioning data markets. For data-driven companies to emerge and flourish a well-functioning data market would be a strong enabler. In such a market, there are both producers of data, refineries known as data factories, and companies turning refined data into products and services.

The construction of a *pan-European data infrastructures*, like the proposed European Data Space, may be an opportunity after taking special attention for its suitability in AI-development and AI-research. A data infrastructure may be considered the best mechanism to build autonomy in AI in Europe as for machine learning the data to a significant extent determine the outcome. For health, but also for all other sorts of relatively rare data resources, coping with data without accumulating them, and respecting privacy, may be key.

2.2.3 Experimentation and deployment in AI

AI-chips and m*icroprocessors* refer to a new generation of microprocessors that are specifically designed to process artificial intelligence tasks faster, using less power. AI chips are expected to play a critical role in economic growth because they will provide advanced AI-processing for mass-market AI apps and move AI capabilities to (increasingly smarter) cars, homes, robots, manufacturing chains, weapons, electronic devices, and all sorts of things connected to the internet.

AI-targeted *high-performance computing* High-performance AI is dependent on high performance computing. However, the requirements for high-performance computing infrastructure in AI could be markedly different from those in traditional areas of high-performance computing. Also, while AI services need to be developed, deployed, and shared across a large variety of computing environments, we assume Europe will establish an increasing number of large-scale AI projects that will need support from larger regional or even European infrastructures.

The *5G-mobile network* is expected to speed up a move of AI computations from the cloud to the edge or IoT devices. Next-generation edge convergence with AI systems on chip: 5G will enable edge devices to seamlessly move between indoor and wide-area environments. These same 5G interfaces will undoubtedly be converged with neural network processing circuitry.

2.4 Possible programs

The PPP will develop a number of co-funded programs. Here follows some examples of possible programs. Ensuring the trustworthiness, correctness, safety, leveraging European strength in automated reasoning. Developing the role of AI in resource allocation, efficiency of services and use of resources, leveraging European strength in AI planning, scheduling, optimisation. Dealing with multiple, conflicting objectives and preferences, between humans as well as between AI agents, leveraging European strength in multi-agent systems. Talent scouting, training, exchange, encouragement in EU across the border of academia – business including start-ups and scale-ups at world level. The accumulation of personal data is undermining proper privacy even if protected. New regulations beyond GDPR need to include technology like differential privacy. Explainable when necessary explain decisions: by example, text, or visualization. Certification and/or watch-dog technology needs to be set up: practical procedures and standards how to do it efficiently in practice.

3. Technology enablers of AI

A weakness in Europe is that the discussion on AI has become very siloed. Contradictions have been created between symbolic AI and machine learned, data-driven, and bottom-up AI, between sensory data knowledge versus real human knowledge, between humane AI and not-so-humane AI. In the end these oppositions are not fruitful nor will survive history. Therefore, Europe should work on obtaining basic and fundamental knowledge in AI without prejudice for the technology. And, Europe could work on the integration of the valuable elements of symbolic and learned knowledge without denominating either technology as humane AI.

3.1 Sensing and Perception in AI

Digital sensory information was used for efficient reproduction until machine learning provided the first application areas where the machine has reached a human-level understanding of what is seen. The same holds for audio and lingual understanding of text. Well trained machines provide a level of advice on dermatologic disorders compatible to the best human performance. The same holds for a broad range of ophthalmic diseases. In manufacturing and agriculture, visual quality control and anomaly detection is on the verge of being standard. New businesses have been built on the automatic reconstruction of 3D space from camera recordings. Diverse activity recognition will find its application in surveillance in a broad range of industries, camera-assisted care, quality inspection, numerical assessment of quantity, and autonomous vehicles.

Geometric Deep Learning is crucial to topics as diverse as the analysis of social interaction and its consequences for the prediction of fashion and societal moods, protein-protein interactions in biochemistry for drug design, bonds between atoms within molecules in material strength, and robustness of planning transportation in road networks. Anything where links between entities is the crucial aspect. Recent years have seen a surge in the deep learning of graph-structured data capturing larger and wider structures in their analyses.

Robot Learning. Effective learning of actions and controls from the right amount of data and instruction for real-world robot aims at significantly improving robustness and flexibility of robots that interact with the real world. Machine learned robots will in the end be able to cope with changing environmental conditions. Not only will such enhance the robustness of the robot but also its precision and its capacity of collaborating with a human user. How should the robot move? How to act? How to interact? The application in manufacturing, agriculture but also in surgery has only just begun.

3.2. Data, knowledge and learning in AI

Machine learning is concerned with algorithms that modify their own behavior based on observations, that can be examples or results of the program’s actions on the world. In the former case, also known as statistical learning, the goal is to build a model of the concept under study from a dataset of examples. Within the ‘learning from examples’ paradigm, one usually distinguishes supervised learning, in which examples are labeled with the target concept, which a category for classification problems, or a continuous function, for regression problems), semi-supervised learning, where usually only a small fraction of the examples is labeled, and unsupervised learning, in which no example is labeled, and classes/outputs need to be identified as well. When the observations are results of the program’s actions, reinforcement learning aims at identifying the best policy choice of action depending on the state of the system) in order to optimize some external reward resulting from its actions. Other systems are for instance recommendation systems.

The most well-known approaches to-date in machine learning, that are responsible for the recent successes of AI are based on neural networks. It has resulted in a world-wide dissemination in computer vision, signal processing, natural language processing, where new application domains appear continuously. Neural networks have revolutionized supervised learning as well as reinforcement learning, because they perform end-to-end learning: not only do they not require any human-designed features to describe the data, but they actually build such representation of the data-oriented toward the task at hand. This allows building powerful generative models known as GANs. Subject of current cutting-edge research: transfer learning, domain adaptation, explainable machine learning and few-shot learning are being addressed.

Although contemporary machine learning algorithms achieve fascinating results even at the borders of human performance, they often remain inefficient, unreliable, brittle, or require manual tuning. Many contemporary machine learning algorithms are still comparably badly understood. As a result, they require manual tuning, can be inflexible and sometimes behave erratically. And there are essentially new elements to the machine learning tree: causal reasoning of learned systems and reinforcement learning where the reward not the label determines the capacity to learn. The development of efficient and reliable learning systems with theoretical guarantees is in order. One particular form of AI Is automation of AI including the choice of algorithms and parameters and configuration in the search community. The main challenges ahead are the identification of descriptive features of datasets in an ML-context or problems in search that would allow instance-based automatic configuration.

3.3. Reasoning and decision-making in AI

Reasoning is any way to infer rational conclusions or making reasonable predictions from available knowledge. Also, for reasoning to be meaningful in addition to correct, it needs to be relevant to any situation at hand. Symbolic reasoning involves the explicit embedding of human knowledge and reasoning into computer programs. Sometimes they are embedded in software systems, sometimes they are components of technologies known as for instance knowledge-based systems, knowledge graphs, ontologies and semantic web, expert systems, and logic programming. Reasoning methods as constraint solving, model checking and automated theorem proving, and methods from SAT and SMT solving, are today widely used for hardware design and verification, software verification in mission-critical systems, planning and scheduling of a number of industrial tasks like robots, air-traffic control, traffic routing, industrial automation. The systems are used to both design by verified synthesis and certify real-world complex systems, both human-designed like air, train and car control systems, stock-trading systems, and automatically designed such as neural networks and other results of machine-learning algorithms used in industry, putting formal guarantees on their use.

Formally verified software and hardware stacks include operating systems, compilers, drivers, industrial encryption schemes. Such components are becoming critical and preventing billion-scale damages resulting from hardware and software bugs, expensive system malfunction as in Mars-landers, and today's wide-scale hacking attacks on practically arbitrary parts of industrial, economic, governmental infrastructure in banks, hospitals, and the military.

Search and optimisation in AI. AI-based search methods include a large variety of techniques to solve tasks and find solutions to optimization problems that are ubiquitous both in industry and public sector. Examples include minimizing energy use, waste, emissions, and cost, drug development, optimizing telecom bands, logistics, supply chains, crew rotation, scheduling, resource use like the use of health care resources during crises, and the use of expensive equipment, like MRI-machines and image and communication satellites. Many of these problems are specializations of research areas in AI, like TSP, quadratic assignment, vehicle routing, as well as their numerous variants on costs and time windows. AI-based search methods pertain to such challenges, which, depending on their nature, are known as combinatorial or continuous optimization problems. AI approaches for combinatorial optimization include constraint programming, AI-planning and scheduling, and are close to reasoning methods. AI-based search methods pertain to such challenges, which are known as combinatorial or continuous optimization problems. Metaheuristics are more general problem-solving techniques that can be adapted and applied to create new market opportunities and handle the increasing complexities of our modern digital world. Several metaheuristics are inspired by nature, in evolutionary computation, simulated annealing, or swarm optimization approaches. They are increasingly applicable and powerful for real-world problems as for instance the ones listed above. These methods are important both for improving global competitive positions for Europe’s highly capitalized industries, increasing efficiency in its public services, and meeting the grand challenges of our time.

3.4. Action and interaction in AI

AI has traditionally focused on full automation, where the computer completely solves a problem without human interaction. However, there is evidence that rather than replacing humans, a better approach is to design systems which allow humans and AI-tools to collaborate effectively. It is also the case that even if a taxi drives completely autonomously, it still needs to interact with the passengers. Effective human-AI collaboration relies on techniques such as natural language understanding, gesture and activity recognition, understanding intention, creating and maintaining shared mental models, and interaction design. This is also an important aspect of maintaining meaningful human control through for example human-in-the-loop, human-on-the-loop and human-in command.

Interactive learning. Machine learning has an excellent capacity of reproduction of the given samples. It does not immediately provide a cause of the automatic conclusion. For a deeper understanding as may be needed in diverse professional services as law and finance, interactive and multi-agent learning and causal modelling are important for acceptance matching the overstressed expectation of robust intelligent behavior attributed to AI. These techniques will be important for high-stake real-world applications.

Human-centric machine learning. As the use of machine learning becomes widespread in human-centric applications and algorithmic decisions are more consequential to individuals and society, key limitations of today's machine learning systems need to be identified and presented in the result. Algorithmic discrimination against minorities, manipulation of human decision-making, spread of misinformation and increase of polarization are simply impossible to tolerate. In the end, it will require close cooperation between regulation and the development of technology to achieve this at the intersection of machine learning, causality, human-computer interface, differential privacy, and computational ethics. Transparency, accountability, interpretability and fairness of the algorithmic decisions, amenability to legal and technical certification, accountability and verifiability are all relevant here.

Explainable AI. Evolving AI-techniques and AI-based systems from black boxes, which they largely are today, to techniques that can complement the decisions they reach with machine- and human-understandable explanations as to why each specific decision was reached, is a necessity for extending the application domains of AI to any domain where machine-human synergy or the accountability of AI are of importance in medical diagnosis, autonomous driving, smart farming, media verification and fact-checking, to name a few. This involves not only focusing on AI-based techniques that can inherently provide some evidence on the decisions they make reasoning systems, or simple versions of deep learning architectures comprising just a handful of layers where some sense can be made as to how the network works, but more importantly focusing on the fundamental issue of how to endow with explainability properties the most complex of the present and future deep learning architectures that can provide state-of the-art results and will be deployed in real-life applications. Explainable AI comes in many shapes and forms: textual explanation, explanation by example, visualization of the decision space, rendition of the decision loop, uncertainty expression.

Robust Machine Learning. As machine learning technologies are progressively deployed across the sciences and into the real world, it is becoming more important that they can reliably perform well, when applied in settings different to those during training. One set of techniques applies adversarial manipulations during training to make the classification robust. Another set of techniques solves learning when only unbalanced, messy or heterogeneous data are available. Applications range from medicine where data are scarce, fault detection in industry where data are intrinsically unbalanced, environmental sciences, in manufacturing for assisted design, the robustness of autonomous vehicles and industrial control to handle also unseen situations.

3.5. Systems, hardware, methods and tools in AI

To develop truly intelligent systems, many different AI-components need to be integrated into working systems with system properties and system guarantees. Developing a science for developing, analyzing, operating, monitoring, maintaining and extending AI systems is therefore greatly needed. This would complement the impressive progress in developing individual AI algorithms and components. This area is closely related to robotics.

On *hardware and software infrastructures*, deep learning comes with a need for GPU-clusters, publicly available software resources such as github and publicly available software suites such as TensorFlow, Caffe and Pytorch. These toolsets have substantially contributed to the current wave of deeply learned AI as they have widely spread modern technology in rapid pace. As to the necessity of developing hardware in Europe, there are two different, almost orthogonal directions: on one side as the size of data will grow, there will be a need for machine learning and large-scale knowledge AI-algorithms to scale up to exascale machines. On the other side, latency issues will make it essential to use edge solutions, which means deploying algorithms on limited computational resources, and adapting algorithms to small sample sizes and domain generalization and federated or privacy-preserved learning.

3.6. Possible programs

The PPP might focus on a number of co-funded trans-sectoral programs in areas such as the following. *Learning reliable AI for banking, medicine, agriculture, professional services, remote sensing, chemistry, social media trend prediction, manufacturing, etc. Learning over multi-modal data (vision, audio, speech, etc.). Geometry for real-estate,. Learning in the context of time series data. Robust AI, including planning and interaction between agents, as one of the most important aspects of trustworthy AI, applies to all application domains. Interactive AI, leveraging humans in the loop, and human-centric settings. AI for earth and space science, sustainability and environment. Automated reasoning for provably ensuring safety and correctness of hard- and software systems in critical applications in areas such as transportation, healthcare and finance. Next-generation robotics, with an emphasis on learning from small and large data, robustness, safety, correctness, efficiency and flexibility, with applications in healthcare as well as industrial and food production. Next-generation transportation systems across all modalities.*

References

[1] “European enterprise survey on the use of technologies based on artificial intelligence”, 2020, the EC.