

Dutch Artificial Intelligence Manifesto

Special Interest Group on Artificial Intelligence¹, The Netherlands

Executive Summary

Artificial Intelligence (AI), the science and engineering that studies and creates intelligent systems, has become a disruptive force revolutionizing fields as diverse as health care, finance, law, insurance, HR, communication, education, energy, transportation, manufacturing, agriculture, and defense.² Driven by the increased availability of compute power, access to massive amounts of data,³ and advanced sensor technology, AI techniques - such as reasoning, imaging processing and machine learning algorithms - have become powerful enablers of automation, predictive analytics, and human-machine interaction. AI has already changed online interactions in the retail sector (e.g., recommender systems and chatbots) but also has enabled sophisticated AI assistants that profoundly affect how people live, work, and play.⁴ To ensure these developments will also boost the Dutch economic system in terms of jobs and GDP and the outcomes will be beneficial for all, we should invest in AI science and technology development, in understanding the socio-economic impact of AI, and in the ethical and legal constraints for AI technology.⁵ The Netherlands is well positioned to contribute to these developments on all three dimensions and to benefit from them as strong enablers are in place including a strong AI knowledge basis as well as strong digital absorption and economic innovation. AI talent is key. To avoid a brain drain, substantial investments are needed in attracting, training and retaining human talent.⁶ In the meantime, a technology race has started with the US taking a leading role, China closely following and heavily investing in AI,⁷ and Europe still in the process of formulating its AI strategy at EU as well as national levels.⁸ We urgently need a national agenda for AI that provides a national strategy that is supported by academy, industry, and government. The Netherlands must make substantial investments in high-quality Dutch AI research and innovation to maintain a leading position in AI and to benefit from its huge potential.

¹ The Special Interest Group of AI, SIGAI, is a special member of [IPN, the ICT Platform Netherlands](#), representing all computing science academic institutes and researchers in the Netherlands that perform AI research. All academic institutes represented have contributed to this document (CWI, LU, RU, RUG, TUD, TUE, UM, UT, UU, UvA, TiU, VU).

² See e.g., [Deloitte in WSJ July 2015](#). Also see <https://aiindex.org/>, a US initiative which reports that the number of AI papers produced each year has increased by more than 9x since 1996, introductory AI class enrollment (at Stanford, but this trend is visible more broadly) has increased 11x since 1996, the number of active US startups developing AI systems in the US has increased 14x since 2000, the annual VC investment into US startups developing AI systems has increased 6x since 2000 (at >3 billion dollar, which represents only “a very small sliver of total investment in AI Research & Development” in the US), and the share of jobs requiring AI skills in the US has grown 4.5x since 2013.

³ Those with access to data have an edge over the competition in the AI era, see [The Economist, May 6, 2017](#).

⁴ See e.g. [Forbes July 2018](#), [Forbes August 2018](#).

⁵ See also the [ASILOMAR AI principles](#).

⁶ See McKinsey Global Institute’s report: [Modeling the impact of AI on the world economy](#) August 2018.

⁷ See McKinsey Global Institute’s report: [Artificial Intelligence: Implications for China](#) April 2017.

⁸ See e.g. the report from [Denkwerk July 2018](#). See also the [Overview of worldwide AI strategies](#) at the national level, and [NRC 25 April 2018](#).

In this manifesto the Special Interest Group on Artificial Intelligence (SIGAI) proposes a research agenda and identifies priorities that require investments to ensure AI research in the Netherlands is able to establish and maintain its leading role in the world. How successful the Netherlands will be will depend on the Dutch government prioritising research in AI. We have defined a research agenda that identifies (1) priorities, (2) the need to invest in AI foundational research, and (3) unique opportunities to invest in cross-cutting, multidisciplinary challenges. Our recommendations are based on input from all AI research institutes in the Netherlands.

The Netherlands stands out internationally in the high quality of its AI educational system but we do not yet sufficiently **exploit our capability to increase our national AI talent pool**. Dutch academia, industry, and government should **create a strong national AI alliance to promote R&D in AI**; and should invest in a **strong AI infrastructure** that is needed to benefit from AI by providing access to quality data for research in, e.g., health.

In order to ensure Dutch AI research can stay competitive world-wide, it is essential to **invest in the foundations of AI research**. We have identified seven foundational AI areas strongly represented in the Netherlands, representing broad areas typically identified within the field of AI:

- Autonomous Agents & Robotics,
- Computer Vision,
- Information Retrieval,
- Knowledge Representation & Reasoning,
- Machine Learning,
- Natural Language Processing, and
- Planning & Search.

Based on the strengths in each of these areas, we have identified fundamental research challenges that will ensure that the Netherlands will continue to be able to conduct cutting-edge AI research that stands out in the world. We focus in particular on how the academic AI community in the Netherlands can contribute and where investments are needed to ensure that the world-renowned Dutch research in AI is strengthened.

AI is having a big impact on the world-wide society, and can be a tremendous opportunity to increase the quality of life of humanity. **AI technology** will provide new opportunities for **innovation** but will also raise several **cross-cutting challenges** that we should address. Because society is relying more and more on decisions that are taken by or together with AI systems, their role and impact on society will increase, and there is a need for AI techniques and models for making these systems socially-aware, explainable, and responsible. These priorities are visualized in a grid (Figure 1) composed of the foundational AI areas and cross-cutting challenges.

Finally, we believe It is important to educate the broader public about current AI driven changes and developments.⁹

⁹ See also [Denkwerk July 2018](#).

Introduction

AI systems are capable of sensing their environment, learn from and reason about it, and change it based on advanced decision making. AI already has had a big impact on our society; this impact will further increase due to ongoing algorithmic developments, the availability of data and increasing computational power, advances in sensor technology and robotics. AI applications are becoming ubiquitous in all areas of our society including science, industry, health, high tech, energy, public safety, food, retail, and education.

In order to enable and facilitate the Dutch AI research community to have a significant impact in these areas and increase its ability to compete internationally, we have formulated a clear research agenda. To this end, we have identified priorities that require investments in Dutch AI research and innovation. First of all, this includes the **general priorities** of investing in AI education, a national AI alliance, and AI infrastructure. Secondly, to ensure that next-generation AI systems are accurate, reliable, and robust, we need to invest in **foundational AI research**. Thirdly, due to the enormous impact AI has on our society, **cross-cutting challenges** related to requirements of AI algorithms are becoming increasingly important. Figure 1 provides an overview of the relation between AI foundations and key cross-cutting challenges of AI in the form of a grid structure.

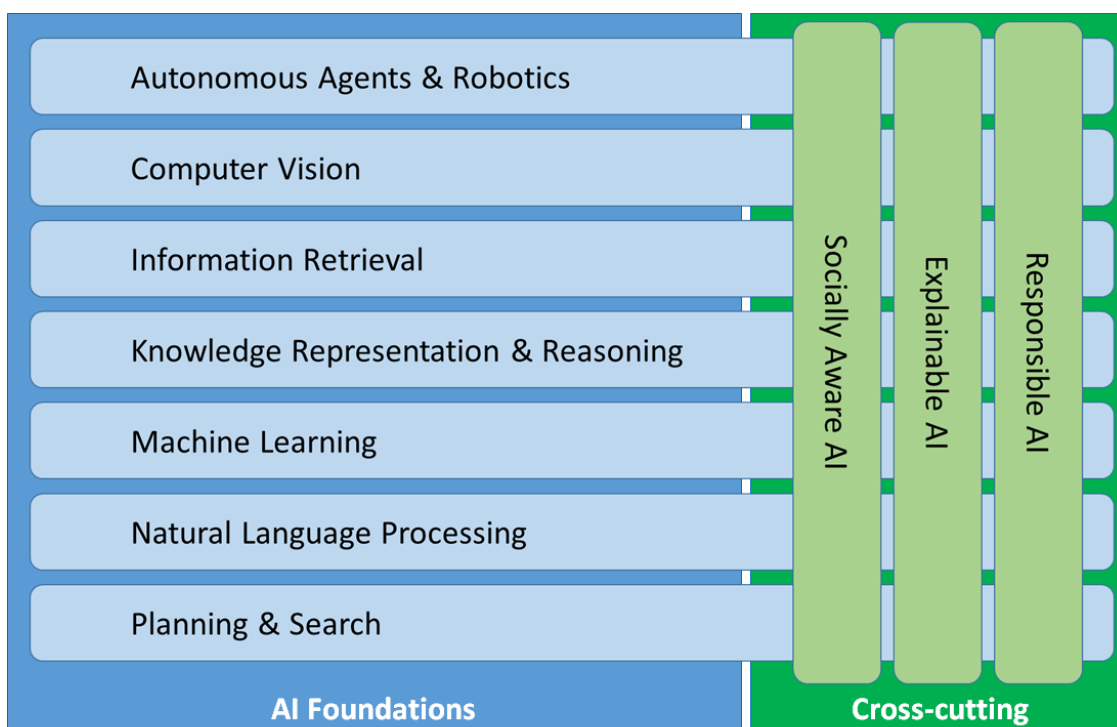


Figure 1: Artificial Intelligence Grid (AI Foundational Areas and Cross-cutting Challenges).

Foundational AI challenges include the development of algorithms that are more powerful, more data effective, more computationally efficient, and more robust. We also believe work is needed on formal and mathematical models of the new techniques that have become available to develop a solid understanding of these techniques (e.g., computational efficiency, convergence properties,

etc.), and tools for verification. We identified seven AI foundational areas¹⁰ in which the Dutch AI community has made (and is expected to make) important contributions:

- **Autonomous Agents & Robotics:** developing autonomous computer systems acting in (either digital or physical) environments in order to achieve their design objectives. Challenges: (1) improving the perception, manipulation, and navigation capabilities of robots, (2) developing sophisticated interaction and coordination models, and (3) integrating different techniques into a coherent decision making architecture.
- **Computer Vision:** obtaining a visual understanding of the world. Challenges: (1) developing novel algorithms for visual interpretation based on precise appearance and geometry understanding, (2) design algorithms which require less expert supervision, and (3) integrating these with techniques from machine learning, natural language, and robotics.
- **Information Retrieval:** technology to connect people to information, e.g., in the form of search engines, recommender systems, or conversational agents. Challenges: (1) algorithmic understanding of information seeking intent, (2) machine interpretation of information interaction behavior, and (3) developing online and offline result generation techniques.
- **Knowledge Representation & Reasoning:** representing information computationally, and processing information in order to solve complex reasoning tasks. Challenges: (1) integrating symbolic with sub-symbolic techniques, (2) robust representation and reasoning techniques for knowledge that is large, dynamic, heterogeneous and distributed, and (3) integration of knowledge representation and reasoning with other AI challenges, such as vision, natural language understanding, question answering, robotics, and others.
- **Machine Learning:** learning from data (using e.g. neural networks also known as ‘deep learning’ and/or statistical techniques). Challenges: (1) integrating pattern recognition techniques with higher level knowledge, (2) applying brain-inspired (spiking) models in energy-lean applications and brain-computer interfaces, and (3) developing uncertainty modelling and handling techniques and more efficient reinforcement learning algorithms.
- **Natural Language Processing:** Extracting information and knowledge about the world from (large amounts of) spoken, written, and signed natural language, enabling human-machine communication, and supporting multilingual human-human communication. Challenges: (1) dealing with the rich variation and cultural differences in language use at personal and group level, (2) achieving technological language- and domain-independence, and (3) achieving naturalness in generated text and speech.
- **Planning & Search:** planning and scheduling, heuristic search and optimization. Challenges: (1) developing uncertainty modelling and handling techniques, (2) developing algorithms to support sequential decision-making under uncertainty and involving multiple parties, and (3) combining reasoning and machine learning algorithms (as in, e.g., AlphaGo).

As a consequence of the rapid developments in these foundational areas, AI is transforming our society by changing the nature of work and the way humans relate to machines. AI should be human-centered: the key to using Artificial Intelligence effectively is to augment and assist humans, not to replace them. Figuring out how humans and AI can collaborate effectively will have large economic and societal impact. It therefore is important to research and develop algorithms that have in-built properties that facilitate collaboration between humans and machines to jointly perform

¹⁰ Based upon among others <https://ai100.stanford.edu/2016-report> and <https://www.ijcai-18.org/cfp/>.

advanced tasks. To fully address this challenge, a fundamental step in AI research is required, that starts from the future human-AI relation. This also introduces new cross-cutting research challenges and opportunities of a multidisciplinary nature. We identify three cross-cutting challenges for sustainable next-generation AI systems to which the Dutch AI community can make strong contributions in the coming decade:

- **Socially-Aware AI:** Next-generation AI systems will become proactive and collaborate in a personalised way with humans. Challenges: design AI systems that (1) allow for an effective social and cognitive interface with humans, (2) are able to interpret, reason about, and influence human behaviour, (3) are able to interact, collaborate and coordinate their behaviour with human beings.
- **Explainable AI:** Next-generation AI systems will perform sophisticated and advanced tasks that were previously performed by humans and moreover their decisions will significantly affect the life of humans. It is therefore of utmost importance that AI systems are not only accurate, but also able to explain how the system came to its decision. Challenges: develop new algorithms, which (1) by design can explain their rationale, (2) do so in an intuitive, human-understandable manner, and (3) explain why their underlying mechanisms produced the AI's behaviour.
- **Responsible AI:** Next-generation AI systems will enable the automation of tasks that humans used to perform and the automated processing of huge amounts of data. This should happen in a responsible manner. Challenges: design AI systems which (1) allow to integrate our moral standards for responsible data processing, (2) integrate our moral, societal and legal values, and (3) verifiably meet our moral standards while processing the abundance of (sensory) information available worldwide.

The theme of human-AI collaboration and the question of how AI can effectively augment and assist humans is also at the heart of the Hybrid Intelligence proposal funded by a Zwaartekracht grant from the Dutch Ministry of Education, Culture and Science in 2019.¹¹ The project aims to address each of the three cross-cutting challenges mentioned above (social awareness is part of the collaboration theme) and adds a fourth theme Adaptive AI.

Finally, the Netherlands can have a high impact when next-generation AI systems are employed in specific domains, in particular in the life sciences & health, high tech, energy, food production, the public sector (e.g. government and law), and education domains.¹²

¹¹ See <https://www.hybrid-intelligence-centre.nl/>.

¹² In a [2017 briefing note](#), McKinsey notes that the “public sector, and healthcare [have] captured less than 30 percent of the potential value” of using AI for data analytics (radical personalization, massive data-integration capabilities, and enhanced decision making can act as disruptive forces in these domains). It is advised to make a more thorough analysis of the [Dutch key sectors](#) and focus on a specific selection, a strategy also proposed in the [Villani report March 2018](#). The current selection is based on high potential (healthcare and high tech), sustainability (energy), leading by example (government), and need to build a larger talent pool (education). Also in line with the Villani report, it is advised to consider *interdisciplinary institutes* for these specific focus areas.

Research and Education Agenda and Priorities

The Dutch AI research community is ideally positioned to play a key role in the development of sustainable next-generation AI systems. The research agenda that we propose starts by identifying a number of important general priorities. After that, the main priorities related to the two dimensions of the AI grid (foundational areas and cross-cutting challenges) are elaborated.

General Priorities

There are a number of general priorities that require investments to ensure the Netherlands can face international competition and maintain a leading position in the academic field. The Netherlands has an excellent educational infrastructure that offers a strong opportunity to invest in talent. To unlock this potential, investments are needed to be able to maximally use and exploit this infrastructure in order to be able to educate larger numbers of AI experts. A similar point can be made with regards to the Dutch ecosystem of academic institutes, companies and government. There is much potential but an investment is needed to shape a coordinated effort to by means of a Dutch AI alliance to identify the needs of the Dutch economy in particular.

Invest in AI curricula to train and sustain an AI-capable talent pool

There is a shortage of skilled AI developers. The skills required for advanced AI system development are relatively rare at present. PhD-level expertise is currently required to understand AI systems and develop novel AI algorithms but current trends such as automated AI (e.g., AutoML) will make these available to trained AI students too. “Nations that develop education, training, and immigration policies to recruit and train top talent – from their country and from others – will have an edge on others”.¹³ Dutch higher education stands out in the world by the relatively large number of high quality AI curricula. Seven universities offer programs on Artificial Intelligence and these programs are highly ranked in the world.¹⁴ However, currently universities are not able to service the much larger numbers of students who want to study AI (>500 applications where resources are typically available for 120) due to resource limitations.¹⁵ Moreover, diversity is an issue in AI as more broadly within other disciplines.¹⁶ **A large investment is needed to enable universities to educate larger numbers of next-generation students who aim to be AI experts and to promote and implement STEM education more broadly, taking also diversity into account. Such an investment can also be used to train and educate company employees in AI.**

¹³ See [CNAS report July 2018](#).

¹⁴ The Netherlands has a relatively high number of institutes that offer AI programs compared to other EU countries, see e.g. for comparison <http://www.aiinternational.org/universities.html>. These programs are ranked 59-195. See also, for example, [Analytics India January 2017](#) which reports that 4 out of the 10 leading Master’s Programs from around the world are Dutch.

¹⁵ See also this [NOS July 2018 report](#). The [Villani report March 2018](#) suggests to at least *triple current numbers*.

¹⁶ See also the [AI4All initiative](#) to train a new, more diverse generation of AI technologists, thinkers, and leaders. The [Villani report March 2018](#) also promotes Inclusive and Diverse AI.

Invest in national AI alliance with academia, industry, and government

Initiatives such as [ICAI](#) and European collaborations such as [CLAIRE](#) and [ELLIS](#) aim at enabling the AI community at national and European levels to enhance their effectiveness and cooperation to do AI research. However, we should also realize that private sector companies are playing a large role in developing and utilizing AI.¹⁷ Many initiatives in the Netherlands are developed bottom-up, where the US and China government initiatives play a large role. In a bottom-up fashion the Dutch robotics community has organized itself in the [Holland Robotics](#) community. AI more broadly has given rise to a plethora of innovation labs in all sectors with less organization and structure so far. This bottom-up organization leads to inefficiencies and does not make use of the potential for cooperation. For the Dutch government to effectively harness AI technology, it will need to stimulate innovation occurring in private companies, to adopt AI technologies itself, and to ensure that the wheel is not reinvented by making sure expertise is shared whenever possible. VNO-NCW and MKB Nederland have also made a recent call to action.¹⁸ **It is important to organize key Dutch national players from academia, industry as well as government in a national AI alliance. This alliance should develop a national agenda for AI & Robotics and invest into public-private cooperation and projects.**

Invest in an AI Infrastructure

“AI will augment the national power of those countries that are able to identify, acquire, and apply large datasets of economic and military importance in order to develop high-performance AI systems.”¹⁹ To this end, we need to make sure that we create access to large quantities of the right type of data. New ways of producing, sharing and governing data by making data a common good are needed.²⁰ Part of the challenge here is how to live by our values and protect e.g. privacy as regulated in GDPR while at the same time make e.g. healthcare data available for next-generation AI systems to benefit from this advanced technology.²¹ **An important challenge is to identify which types of data are crucial at the Dutch national level (our healthcare but also scientific data sets are good examples) and how we can make this data accessible to exploit the benefits that AI can offer.**²²

AI Foundational Priorities

Autonomous Agents & Robotics

While AI traditionally focuses on specific capabilities such as natural language processing, planning, machine learning, search, and vision, the integration of various capabilities is becoming the main research focus of the autonomous agent research field. An autonomous agent is capable of sensing its environment, reasoning to decide on a next action, and performing actions that maximise the chance to successfully achieve their objectives. Moreover, the interaction between autonomous agents and the corresponding notion of social intelligence have become the focus of the multiagent systems research field. A key issue here is how to predict and control emergent behavior in complex

¹⁷ See [CNAS report July 2018](#).

¹⁸ See this [NOS July 2018 report](#).

¹⁹ See [CNAS report July 2018](#).

²⁰ [Villani report March 2018](#).

²¹ See [Centre for Data Innovation March 2018](#).

²² It is advised to further stimulate and invest in initiatives such as <https://researchdata.nl/>.

agent systems. Europe, with the Netherlands playing an especially large role, has been a main contributor to the worldwide research on autonomous agents and multiagent systems. The community of autonomous agents and multiagent systems in the Netherlands has strong researchers in Amsterdam, Leiden, Delft, Groningen, Maastricht and Utrecht. Key contributions have been made to engineering autonomous agents and multiagent systems, logics for multiagent systems, argumentation and dialogue systems, virtual and emotional agents, decision and game theory, negotiation and teamwork, self-organisation and swarm intelligence, normative multi-agent systems, trust and reputation, and social systems with applications in among others virtual reality, gaming, simulation, supply-chain management, traffic and transportation, energy grids, healthcare and other branches of industry. **The main challenge for autonomous agents research is to integrate AI techniques into a coherent decision making architecture, and to develop interaction and coordination models and techniques.**

The Netherlands has been a key player and actively participating in the worldwide robotics community, both from a technical but also from an AI and human-robot interaction perspective. The academic robotics community has always had a strong focus on a human-centered approach to robotics. However, whereas the Dutch industry has organized itself in Holland Robotics, organizing the academic community and connecting it better with industry requires an investment to ensure innovations are driven by both. **The main challenge for robotics is to further improve the robot's perception, manipulation, and navigation capabilities and research and develop sophisticated cognition and collaboration capabilities so robots can interact socially and cooperate with humans.**

The use of advanced machine learning techniques on robots require many trials, which are often only feasible when performed in a simulation of the robot. A key challenge is to research the interaction between the simulation and the robot. How should we deal with the “reality gap”, i.e. differences between simulation and real-world physical robotics. How can we improve simulators using data from robots, how can we learn to choose between data from the simulator (cheap) or real robot (expensive), and how can we learn the best policy from the resulting dataset?

Computer Vision

Computer Vision, and more generally Intelligent Sensing²³, is of central importance to many fields, of which health and transportation may be the two most visible. The world is adapting swiftly to visual computing, communication and intelligence via Internet, mobile phones and other with camera-equipped devices. The inevitable future is that the story of our life, and that of a billion others, will be told by video at home, at work, and on the road. Making sense of visual data offers a wealth of opportunities for science, for health, for public safety, for well-being, for transportation, and for businesses. However, at the same time automatically understanding the full complexity of visual content presents one of the grand challenges in computer science, covering topics such as color processing, semantic understanding, 3D reconstruction, interactive picture analysis, image and video retrieval, human-behavior analysis, and event recognition required to be used as vision system for e.g. robots and autonomous driving systems.

The field is at the forefront of the current revolution of AI research. The superhuman image classification performance by deep learning in the ImageNet competition is the leading example to

²³ Note that audio processing is covered by the NLP core technology.

stress the breakthrough of AI in general. The field's major conference, the Conference on Computer Vision and Pattern Recognition is the highest ranked peer-reviewed publication in all of computer science. The Netherlands has a long tradition in the field with strong groups in broad and technical universities across the country. They participate in the research school for computing and imaging and meet at the yearly Netherlands Conference on Computer Vision. The Netherlands also plays a role of international importance, as evidenced by the University of Amsterdam's victory in the ImageNet competition, Leiden University having developed the best edge detector, as well as several other image and video recognition benchmarks, the organization of the European Conference on Computer Vision in Amsterdam in 2016, the establishment of two public-private research labs on computer vision by learning with Qualcomm and Bosch and several successful university spin-offs, including Kepler Vision, EUVision (acquired by QualComm), SightCorp, 3D Universum, and Aiir Innovations.

Three challenges in computer vision research are apparent. First, the analysis of images and videos in terms of appearance and geometry have to become increasingly precise and should shift towards interpretation. Second, the integration of core computer vision with machine learning, natural language, and robotics should prosper. Third, the amount of required expert supervision should be reduced, requiring novel forms of automated visual analysis. Together, solutions to these challenges will enable the automated understanding of videos in terms of interaction, behavior, context and causality. As such, automated computer vision will provide the cornerstone for comprehensive AI systems, used to assist cross-cutting research in socially-aware, explainable, and responsible AI. At the same time, there is a technology push by the Internet of Things causing massive amounts of video cameras to become tiny, connected and live-streaming 24/7. Body worn cameras, drones and service robots are just three examples. They may even capture video data beyond the visual spectrum. This is an invitation to move away from traditional computer vision domains, such as consumer photos, television archives and social media. And, instead, emphasize hitherto non-mainstream video domains like surveillance, healthcare and robotics where viewpoints are new, labelled examples are scarce and real-time spatiotemporal 3D understanding is crucial. **The key challenge here is to develop precise automated visual understanding, which is integrated with machine learning, natural language, and robotics, uses less expert supervision, and which generalizes to novel visual domains.**

Information Retrieval

Information retrieval, in the form of search engines, recommender systems, or conversational agents, is the most prominent manifestation of AI in practice today. Information seeking is a key aspect of the human condition and the drive to create technology that helps us access information is as old as the drive to create technology itself. Modern information retrieval technology is centered around interactive methods that learn to assess and improve their results while interacting with their human users. Key steps in these algorithmic processes are obtaining a computational understanding of people's information seeking intent, enabling machines to interpret and assess human information interaction behavior, and developing online and offline result generation techniques that learn how to meet the user's intent and how to improve their results based on observed human behavior.

In terms of obtaining a better computational understanding of people's information seeking intent, approaches so far have mostly focused on static snapshots of users' state of knowledge, either

implicitly by mining interaction logs or explicitly by modeling domain-specific intents. **The key challenge here is to integrate these implicit and explicit intent representations and track and predict shifts in intent as users interact with information retrieval technology, especially in data sparse environments.**

As to improving the ability of information retrieval algorithms to assess the success or failure of their actions based on observing human information interaction behavior, the community has made significant improvements recently through the introduction of new online and counterfactual evaluation methods. **The next big challenge here is how we can setup reliable simulation methods that enable us to evaluate personalized interactive systems.**

Knowledge Representation & Reasoning

Symbolic AI has been an important branch of study within AI. Nowadays symbolic AI and logical methods are very important given the need for explainable and responsible AI, as both require reasoning about AI systems. The Netherlands has been a strong contributor in this field with key contributions in (i) investigations into agent concepts such as knowledge, belief, time, intentions, goals, emotions, and norms, (ii) argumentation theory, and (iii) knowledge representation and reasoning, ontologies, and the semantic web. Research projects aim to understand these notions better and to make them amenable to computerised treatment and processing.

It is a significant challenge how data-driven methods can be connected to knowledge-based approaches. Research developing such connections requires innovative foundations (for instance explaining how logical and probabilistic knowledge and reasoning can be integrated, new algorithms (in particular allowing scalability) and the development and validation of corresponding design paradigms. By connecting data-driven and knowledge based methods, progress can be made in several challenges. First such connections contribute to addressing the long-standing knowledge acquisition bottleneck. Second they contribute to explainable AI by bridging the transparency of knowledge-based methods with data-driven scalability. Third, they can support a responsible AI as knowledge-based methods are more amenable to normative guidance than data-driven methods. **The key challenge is how to integrate symbolic and subsymbolic techniques in AI, which would establish a major breakthrough in AI.**

Recently, probabilistic modeling methods (in particular Bayesian networks) are gaining momentum in knowledge representation and reasoning with uncertainty, for instance in domains such as law and medicine where complex knowledge structures are combined with statistical analyses. Also in argumentation research, originally primarily qualitative and knowledge-based in nature, there are early attempts to make connections to probabilistic and other numeric approaches (such as expected utility decision making). **The key challenge is how to connect data-driven, often numerical methods with knowledge-based, often qualitative methods, requiring new foundational, computational and design research.**

Machine Learning

Recent development in machine learning, and in particular deep learning and reinforcement learning, have fueled enormous progress and excitement in the field of AI. Deep learning is transforming both industry as well as science at a rapid pace due to improved image and audio

analysis tools, better natural language analysis methods such as machine translation, better planning algorithms that have beaten the world champion at Go and so on. Yet, it behoves us to acknowledge that there are many things we cannot do very well yet. For instance, while deep learning is very good at pattern recognition ('fast thinking'), it is not yet very good at high level reasoning ('slow thinking'). Our models simply do not understand the world yet at a level that humans do, and this lack of context makes it very hard for them to generalize to new situations at the same level as humans do. For example, humans can effortlessly generalize from a single example, where ML algorithms horribly fail. Humans can easily derive cause and effect by simply observing the data, ML algorithms do not have that level of understanding. Machine learning has not been able to perform at the same level as humans in question-answering tasks which require high level reasoning. Upon closer inspection, humans only achieve good performance at such tasks within domains in which they have large amounts of reliable background knowledge and understanding; one of the main challenges therefore is to integrate such knowledge into machine learning algorithms. These limitations point to interesting directions to extend machine learning: incorporating reasoning, heuristic search, or causality and the laws of physics in our models, and incorporating the knowledge of massive relational databases into our models. Such extensions may also contribute to addressing the problem that machine learning models are black box models which are not transparent to a human. problem Also, making ML more data and compute efficient, perhaps by finding inspiration in natural intelligence, will constitute important progress. DeepMind's AlphaGo is a good example of the breakthrough that can be achieved when heuristic (Monte-Carlo) search and deep (reinforcement) learning are combined successfully. In the Netherlands, we have the expertise to bring these fields together and build a more comprehensive ML framework. Such a framework should also include techniques for parameter and structural optimization of machine learning algorithms. **An important challenge in ML is to integrate pattern recognition techniques with higher level knowledge such as reasoning, causality, knowledge graphs and simulators that model physical constraints.**

Many situations where AI will be applied in the near future, such as control of autonomous vehicles of smart electricity grids, do not comply with the prevalent paradigm of supervised machine learning. Notably, the ideal output (response) of the system might not be known in all cases. Instead a delayed, noisy reward is provided that serves as a learning signal. As such, data acquisition now depends on the system's own behavior, giving rise to exploration/exploitation trade-offs and the danger of self-fulfilling prophecies. Such problems can be modeled as **reinforcement learning problems**. These problems pose several related challenges: First, data is typically only valid close to current system behavior. Often offline learning using simulations are used to train models (see below for online learning). Thus, as the system is learning data needs to be collected periodically. Learning rich behavior from small datasets is an important research challenge, that might involve enriching this data set using simulated or virtual data. Second, to learn about regimes beyond current system behavior, data also needs to be gathered using a variety of strategies besides the best known strategy so far. Finding good 'exploration' strategies is a defining challenge in reinforcement learning that might require optimizing such strategies in a data-driven way, while constraining them to satisfy constraints on safety, fairness, and explainability. Third, integrating models into reinforcement learning is still a challenge. Finally, reinforcement learning is currently compute and data inefficient. We need better learning algorithms that more efficiently attribute reward to actions. **A key challenge in deep reinforcement learning is to make it more data and computationally efficient.**

Adaptive AI systems can adjust their behavior based on the experience they gain over time. **Developing adaptive AI systems that learn from the data that is generated by the system while it is interacting with its environment is an important challenge that requires on-line learning in real time.** Arguably, digital as well as physical agents (smartphone apps, robots, self-driving cars, smart homes, etc) must become adaptive, (self-)learning systems to optimize themselves on the job, to offer personalization and customization to their users, and to cope with unforeseen and/or changing circumstances. Techniques from evolutionary computing and other biology and nature-inspired techniques may be useful in this (as well as other) areas.

In recent years we have witnessed the emergence of **brain-inspired** (spiking) network models as well as brain-inspired (neuromorphic) hardware such as Intel's Loihi chip. Event-driven models and non-Von Neumann hardware architectures potentially allow for energy-lean computations, but require novel learning methods and programming paradigms before this potential is fully unlocked. Key challenges here are to understand and apply non-traditional ways of representing and reasoning information in neuromorphic computational models, develop energy-efficient algorithms, and apply such algorithms with event-driven sensors and actuators, for example in the context of brain-computer interfaces.

A challenge that directly relates to responsible AI is **uncertainty modelling and handling**. Current supervised (deep) learning methods, for example, often provide no explicit quantification of their own uncertainty at all; Bayesian supervised and reinforcement learning methods do, but their quantification is often far too optimistic: they "think" they perform better than they actually do. The challenge here is to develop novel statistical techniques for reliable uncertainty modelling and handling and incorporate them into deep learning algorithms.

Natural Language Processing

The Dutch AI community includes a number of very strong **NLP** or more generally **computational linguistics research** (CL) groups, some of which operate as applied technology groups in faculties of arts and humanities, keeping a direct link to the field of linguistics and communication and information sciences.²⁴ Since the 1970s, Dutch computational linguistics has played a key role internationally, in machine translation, in parsing and dialogue systems, and have been pivotal in introducing machine learning into computational linguistics. The community cannot be seen separate from the speech technology community, working on automatic speech recognition and synthesis, and the information retrieval and text mining community, with which it shares people, models and history (e.g. statistical language modelling was popularised in information retrieval through cross-pollination with Dutch CL researchers). The Netherlands has been at the forefront of speech synthesis research in the 1980s and 1990s, and of building applications requiring automatic speech recognition technology. **Key challenges are how to handle and exploit the rich variation and cultural differences in language use and communication at both the personal and group level, achieving**

²⁴ The following groups illustrate this strength: the Centre for Language and Speech Technology at the Faculty of Arts at Radboud University, the Computational Lexicology group at the Department of Language, Literature and Communication at VU University Amsterdam, the Language & Computation group at the Institute for Logic, Language and Computation at the University of Amsterdam, the Computational Linguistics group at the Faculty of Arts at the University of Groningen, the research group Language, Logic and Information at the Utrecht Institute of Linguistics OTS, and the computational linguistics group at the Cognitive Science & AI department at Tilburg University, and a special chair on Text Mining (UM).

technological language- and domain-independence, and achieving naturalness in generated text and speech. Areas of research that should be mentioned here include machine translation, natural language understanding and generation, question answering, and, for example, speech related sentiment analysis. Research is not only needed on speech processing but techniques are also needed that are able to process at the syntactic, semantic, and pragmatic level (e.g., integrating insights from conversation analysis).

Development of conversational agents requires new research (1) mixed initiative search, where systems autonomously elicit from users, (2) user modeling, (3) automatically generating information objects and descriptions of retrieval results, (4) stateful search, (5) dialog systems capable of sustained multi-turn conversations, and (6) evaluation. Development of effective online learning and evaluation methods requires (1) dedicated learning methods that are effective in extremely large, non-stationary state and action spaces, (2) unsupervised objective functions for multi-turn retrieval, and (3) mixed methods approaches to retrieval and evaluation based on online, counterfactual and simulation-based algorithms. Finally, speech interfaces are needed, for which an open source, adaptive, language-independent speech recognition system is required.

Planning & Search

AI planning and search aims to provide algorithmic solutions for both single and multi-agent planning and scheduling problems, as well as for search and optimization problems. This involves algorithms to efficiently search in general solution spaces, and methods specifically for planning and scheduling, reasoning under uncertainty, adaptive strategies, and constraint satisfaction. Historically, this is a field of AI where significant progress has come from the Netherlands, because of its strong base in Mathematics and Operations Research. AI for responsible decision making requires thinking through the consequences of decisions into the future. A key issue here, just like in machine learning, is reliable uncertainty modelling and handling: a good decision-maker, whether a human, a machine or a machine-assisted human, needs to have a good idea about the potential suboptimality of its decisions, for example, statistically valid bounds of the probability of bad outcomes. The Netherlands has strong groups in multi-objective decision analysis and optimization, both fundamental and applied.

Responsible decisions also need to be socially-aware, taking into account the involved people and their preferences, and the (reasons for these) decisions need to be explainable to human experts and policy makers and be based on available data. Decision-making must be adaptive, as circumstances in which decisions must be made may change rapidly; moreover, adaptiveness is a necessary requirement for an automated decision-making system to be self-correcting. Although there has been steady progress on algorithms to support automated decision making, **including reasoning about sequential decision-making under uncertainty and decisions involving multiple parties, making these systems more socially-aware, explainable, and responsible is an important challenge.**

Cross-cutting Challenges

Because AI systems will have a significant impact on our society, changing the way we work and the way humans relate to machines. AI should be human-centered: the key to using Artificial Intelligence effectively is to augment and assist humans, not to replace them. Figuring out how humans and AI

can collaborate effectively will have large economic and societal impact. It therefore is important to research and develop algorithms that have in-built properties that facilitate collaboration between humans and machines to jointly perform advanced tasks. To fully address this challenge, a fundamental step in AI research is required, that starts from the future human-AI relation. This also introduces new cross-cutting research challenges and opportunities of a multidisciplinary nature. We identify three cross-cutting challenges for sustainable next-generation AI systems to which the Dutch AI community can make strong contributions in the coming decade: These AI systems should be **socially-aware** to support collaboration with humans, **explainable** to be transparent, and **responsible** so we can account for the decisions made.

Socially-Aware AI

Technological developments over the past decades are reshaping artificial intelligence in various ways, both concerning environments in which intelligent systems are being deployed and concerning environments in which data is being collected. In particular, sustainable next-generation AI systems should be able to understand and reason about their social context, allowing them to engage in sustainable long-term interactions and to interact and collaborate with human beings in order to achieve joint goals. These systems, moreover, should be able to estimate user state (e.g., what does a user know and/or remember) and take such states into account to effectively support their users and personalize the interaction. Examples of systems that require this 'social' ability include smart environments (cities, buildings, rooms), social robots, wearable devices (e.g., for health tracking) and a range of handheld devices. Socially aware AI aims to leverage large-scale, dynamic, continuous, and real-time sensory data as well as computational models of physiological and cognitive processes to recognize individual behavior, discover group interaction patterns, and support human collaboration. Since human communication takes place primarily via speech, automatic speech recognition plays an important role in building socially-aware AI. Moreover as emotions, values, and (social) norms are fundamental in social interaction, this will also require the modelling and use of artificial emotions, values, and norms to enable AI systems to detect and generate appropriate responses. **The main challenge is how to research and develop proactive and personalised AI systems that are able to collaborate with humans taking their emotions, values, and norms into account.**

Explainable AI

As AI systems such as robots and machine learning and decision making algorithms will significantly affect their users, it is important to be able to explain how and why an AI system produced the effect that it did. In many practical cases (e.g., healthcare) it matters how the results of a decision came about and due to the increased complexity of AI systems some form of explanation from such systems will be required. It is a well-known hurdle for today's data-driven artificial intelligence that the algorithms produced behave largely as black boxes. For instance, algorithms trained by extensive data analysis using state-of-the-art deep learning techniques perform well in terms of the input-output function they represent, but it remains hard to make sense of the internal structure of the learnt algorithms. These systems are often aimed at computing optimal decisions and one issue is how to explain the result of such optimization techniques to users. In order to make progress, we believe that there is a need for investing in explainable AI, i.e., in data analysis and machine learning techniques that develop an understanding of the available data, e.g., in the form of complex knowledge structures. The aim is to design tools to help make the inner workings of AI systems more transparent. This requires research in models that are more open to explanation, techniques and

models for generating satisfactory explanations, and intelligible user interfaces for interacting with human users, taking into account that the same data should be explained to different stakeholders differently. **The main challenge is how to develop advanced AI systems which can explain their rationale for how they perform sophisticated tasks.**

Responsible AI

Today's complex AI systems have become less predictable while at the same time their potential impact on our society has increased. It therefore is essential to be able to explore and assess the real-world impact of AI and to help society to anticipate and control the effects of AI.²⁵ In order to be able to accept and trust these systems we will need tools and techniques to make sure these systems will comply with our norms and standards and their operation is sufficiently transparent. Data-driven artificial intelligence techniques are designed for, hence good at discovering patterns in the data.

For responsible decision making, one needs to know how confident one can be that the patterns found will also occur in future data. Current AI systems are either not capable of giving such uncertainty quantification or can be severely overconfident. While this problem may to some extent be mitigated by scaling up existing statistical techniques, there is another issue that requires more radical innovation: because of the aim of objective data analysis, both wanted and unwanted patterns are discovered. For instance, machine learning algorithms run the risk of adopting the same racial, gender or other biases that humans have (e.g., 'most nurses are female'). Even when such patterns are present in the available data, the consequences may have critical consequences for society. What is needed is data analysis technology that can be steered normatively, so that the outcomes meet ethical criteria. These techniques should e.g. comply with the FACT principle.²⁶ This requires the balancing of objective, descriptive aims and subjective, normative aims (cf. what is common in social sciences such as the law). Great impact can be expected from such ethical systems design methods, also with the perspective of ever more autonomous weaponry and privacy-invading investigative measures. New frameworks are needed that can guide us in identifying the moral issues that arise due to the application of AI and help us determine the parameters that we as a society want to be optimized in these systems. In order to achieve this we need precise formal accounts of the moral issues involved and techniques that support verifying that the behaviour of an AI system stays within those parameters. These techniques may also provide an alternative when an AI system is not capable of yielding explanations. Finally, we also need techniques to guarantee that a system respects the user's autonomy, privacy, safety and rights. **The main challenge is how to process the abundance of available data in an efficient and sustainable manner and how to develop AI systems which comply with our moral values to ensure responsible data processing.**

Conclusion

Artificial Intelligence has become a disruptive force revolutionizing many fields in our economy and society. This AI Manifesto has demonstrated the strength and potential of Dutch AI research but also identifies the need for a national strategic AI agenda to be able to compete. If provided with adequate resources, the outstanding and excellent Dutch AI educational system can provide the AI talent we need. By creating a Dutch AI alliance, industry, academia, and government can stimulate

²⁵ See e.g. the [Ethics & Society initiative](#) of Deepmind.

²⁶ FACT refers to questions related to Fairness, Accuracy, Confidentiality and Transparency (see <http://www.responsibledata-science.org/>).

investments in AI nationally. By investing in AI infrastructure, the required resources for AI research and development should be made available. Challenges in seven key AI foundational research areas have been identified that will result in major breakthroughs. The Dutch AI research community is well-positioned to contribute to these challenges but faces increasing competition internationally. Finally, three cross-cutting challenges were identified that need to be addressed to ensure our society is able to handle the changes brought about by next-generation AI systems.