1. Future of machine learning and artificial intelligence

*By Rüdiger **Urbanke** (Professor for Communication Theory, EPFL) with contributions from Emmanuel **Abbe** (Professor for Mathematical Data Sciences, EPFL), Samy **Bengio** (Google Research), Antoine **Bosselut** (Stanford AI Laboratory), Jennifer **Chayes** (Associate Provost of the Division of Computing, Data Science, and Society, Dean of the School of Information, Berkeley), John **Platt** (Google Research), Shai **Shalev-Shwartz** (Hebrew University Jerusalem), Bin **Yu** (Professor for Statics and Computer Sciences, Berkeley), Yi **Zheng** (Institute of Automation, Chinese Academy of Sciences) and edited by Daniel **Saraga***

Key Concepts

Artificial Intelligence (AI) aims to build machines that are able to behave in ways we associate with humans: perceiving and analyzing our environment, taking decisions, communicating, learning, etc**. Machine Learning** (ML) is one approach to AI, developing algorithms that are able to learn autonomously from data. ML uses statistical models, as well as deep neural-network architectures loosely inspired by simple brain models. There are other possible approaches to AI, e.g., in the second half of the 20th century, AI was based mostly on **expert systems**, using rules defined by hand.

Machine-learning algorithms autonomously learn to accomplish tasks by following three broad methods:

- In **supervised learning,** algorithms learn the ability to make the correct associations between a given input and the desired output. They do so by learning on training sets that comprise many correct input/output pairs. A typical example is image **classification**, where algorithms have to put each image in the appropriate category, e.g., cars, trucks, bikes and pedestrians.
- In **unsupervised learning**, algorithms train only on input data. Their task is to uncover patterns in given datasets. A typical example is **clustering**, where algorithms sort a set of inputs into groups that share some similarity, for instance, different groups of customers.
- In **reinforcement learning**, the algorithm repeatedly chooses from a given set of actions in order to maximize a reward function that should lead it to the desired result. Each choice of action enables the **exploration** of the environment (for the long-term reward), as well as its **exploitation** (for short-term reward). A typical example is learning to play a game such as Go, chess or video games, where the reward function increases the score or winning the game. Reinforcement learning is considered to be a promising strategy for addressing complex real-world problems.

These methods use various statistical techniques. Perhaps the most important is based on **artificial neural-networks** that are inspired by rudimentary models of the brain. **Deep learning** refers to models with many layers of neurons.

The concepts of **Artificial General Intelligence** (AGI), Human-Level Artificial Intelligence or Strong AI refer to systems that would exhibit the broad range of human intelligence. They would be able to deal with complex, dynamic, and open environments, which comprises interacting entities, and to perform tasks that are complex and diverse. These systems would be able to learn autonomously in order to perform new tasks.

Scientific Anticipatory Brief abstract

Current status of research

Artificial intelligence (AI) has seen impressive recent progress. This so-called “second wave’’ of AI is made possible by breakthroughs in machine learning, in particular deep neural architectures, as well as an exponential increase in available computational power and access to massive data sets. This trend is expected to continue in the next 5 years, bringing benefits to more and more specialized application areas.

While the progress is impressive, current machine learning systems have significant limitations**. Training the algorithms requires substantial computational power and access to vast datasets, thus allowing only a few organizations to pursue the most ambitious developments.** Machine learning algorithms do not work well for categorical (rather than continuous) variables, or when inputs are heterogenous. The output of the algorithms is brittle by nature and can easily be fooled by specially engineered inputs. **Successful models are nearly impossible to interpret or understand, making their certification a challenge**. The transfer of “acquired knowledge’’ to other problems is difficult, limiting our ability to attack the “long tail of everyday tasks’’. Finally, the current machine learning paradigm is not well positioned to deal with the many types of problems encountered in everyday life, where the environment is highly dynamic and unpredictable, with multiple objectives that might be ambiguous or contradictory, and are embedded in an implicit system of human and social values.

Trends at 10 years - integrating contextual Information and common-sense knowledge

Looking into the future, **the “third wave’’ of AI will consist of integrating contextual information, common sense knowledge and high-order reasoning.** This will enable machines to learn from much smaller datasets than is currently possible, substantially increasing their applicability to a much larger and more diversified set of real-world problems. **They will understand and perceive the world on their own and will be able to perform basic forms of reasoning. They will be deployed much more broadly, increasing the number and depth of human-machine collaborations.** The field of AI should benefit from neuromorphic chips that directly implement neural networks in the hardware and, possibly, from advances in biocomputing and quantum computing.

Trends at 25 years - towards human-level artificial intelligence

The “fourth wave” will involve the development of truly intelligent machines and, possibly, reach Artificial General Intelligence (AGI) where **machines have the capacity to learn any task as well as or better than humans.** AGI will have fundamental implications ranging from our understanding of basic science questions to new applications in virtually all areas of human activity. A survey conducted by the Future of Humanity Institute in 2013 revealed that experts working in the field estimated that AGI had a 10% chance to happen before 2022, and 50% before 2040.

Potential consequences on people, society and the planet

Powered by machine learning and digitization, AI has the potential to impact most human activities and societal issues. It will be deeply embedded in our surroundings, changing the way we live and work. It will transform most industries, such as health care, energy, transport and infrastructures, and will support lifelong education, accelerate scientific discoveries, and impact defence. It will also play a major role in addressing significant challenges such as climate change and a sustainable economy. Human-machine collaborations enabled by AI will become more widespread, impacting a large number of jobs and blurring the difference between humans and machines.

With such huge potential to impact in all areas of society, AI is increasingly considered a strategic sovereign technology. Due to the substantial resources required to pursue cutting edge developments in AI, its advances are mainly driven by large private and public actors. For these two reasons, AI raises fundamental issues about democratic control and is becoming a global geopolitical question.

Detailed table overview of trends at 5, 10 and 25 years

1. Future of machine learning and artificial intelligence
Example of breakthroughs

5 years

Second Wave of AI - statistical ML

- Numerous breakthroughs due to deep learning in various areas (supervised, unsupervised & reinforcement learning).
- Performance gains linked to the exponential growth in computing power and access to data.
- Deployment of AI systems in real-world applications expands, in particular in banking, healthcare, industry and transport. Machine-learning techniques become pervasive and thereby less visible
- The field of **explainable AI** sees a rapid development. This increases reliability, transparency and explainability and will have an important impact effect on applications, as it will then be possible to deploy AI techniques in sensitive domains where liability is paramount.
- In the area of reinforcement learning, important work continues around the implement ethical considerations in the algorithms by rewarding outcomes that fits **human values.**

Bottlenecks & challenges

- Fairness and bias
- Transparency, explainability and trust
- Robustness against attacks
- Transfer learning and complex multi-dimensional objectives

10 years

Third Wave of AI - contextual adaptation

- Integration of common-sense knowledge, physical rules and deductive reasoning.
- As most of daily situations and real-world problems involve small amounts of data and ill-defined goals, the ability to learn from few data points and to deal with open-ended questions vastly increases the relevance and applicability of AI. **This in turns induces an exponential growth of AI knowledge and increases the opportunities for human-machine collaborations, including the augmentation of human capabilities through AI**
- The shift away from huge data sets and brute-force approaches creates incentives for alternatives such as **quantum machine learning**, **neuromorphic computing** (chips mimicking neural networks, directly into the hardware) and **biocomputing** (information processing based on biochemical components, such as nerve cells, DNA, or metabolic processes in the cell, and which that takes advantage of naturally occurring stochasticity and evolutionary processes to treat information). Hybrid architecture combining these approaches with traditional ML might yield unexpected advantages and have strong effects on other fields (for ex. simulating complex gene pathways for gene-editing applications)

Bottlenecks & challenges

- Competing paradigm to the data/computational power advantage?
- Transferring knowledge in different tasks in the real physical world.
- Development of hybrid architectures combining deep learning, quantum ML and biocomputing

25 years

Fourth Wave of AI - human level intelligence

- Brain implants coupled to with AI systems accelerates the development of brain-machine interfaces. These are useful in therapeutic setting (e.g., for neuro-prosthetics) but also open avenues towards augmenting human abilities. It enables discoveries in neurosciences, which brings new insights into human consciousness.
- Progress towards **Artificial General Intelligence** and ability able to deal with complex, dynamic and open environments

Milestones to reach for human-level intelligent systems

1. Answering open-ended questions about a video sequences, e.g. related to people's motivations or hypothetical scenarios
2. Transferring knowledge between different tasks in a creative, innovative and value-creating manner
3. Incorporating emotions and subtle cues necessary for enriching human-machines interactions; understand human values and apply human ethics in a reliable manner,

Challenges

- Dealing with profound societal challenges brought by AGI (e.g. Bostrom, further division of society between with and without access to AGI, geopolitics, reshaping work, etc.)
- Fundamental questions around conscious and truly intelligent machines