European Parliament

Briefing

LEGAL AFFAIRS

Artificial Intelligence: Potential Benefits and Ethical Considerations

KEY FINDINGS

- The ability of AI systems to transform vast amounts of complex, ambiguous information into insight has the potential to reveal long-held secrets and help solve some of the world's most enduring problems.
- However, like all powerful technologies, great care must be taken in its development and deployment. To reap the societal benefits of AI systems, we will first need to trust them and make sure that they follow the same ethical principles, moral values, professional codes, and social norms that we humans would follow in the same scenario. Research and educational efforts, as well as carefully designed regulations, must be put in place to achieve this goal.
- International Business Machines Corporation (IBM) is actively engaged, both internally as well as with its collaborators and competitors, in global discussions about how to make AI ethical and as beneficial as possible for people as society.

1. WHAT IS ARTIFICIAL INTELLIGENCE?

The term "artificial intelligence" (AI) has been mentioned for the first time in 1956 by John McCarthy during a conference where several scientists decided to meet to see if machines could be made intelligent. Since then, AI is usually defined as the **capability of a computer program to perform tasks or reasoning processes** that we usually associate to intelligence in a human being. Often it has to do with the ability to make a good decision even when there is uncertainty or vagueness, or too much information to handle.

As an example, playing chess well, or some complex card games, is believed to need some form of intelligence in a human being, as well as choosing the best diagnosis in a difficult medical case, or creating something new, such as a mathematical theorem or even some form of art, or even driving a car in the middle of a crowded city.

It is clear that this is a strange definition, because it depends on what we consider being intelligent in the behaviour of a human being at a certain point in time. If our belief about human intelligence changes, and we don't believe any longer that a certain task requires intelligence, then a computer program performing that task is no longer part of AI, it becomes just another boring computer program.

The term "artificial intelligence" brings to mind to the notion of replacing human intelligence with something synthetic. At IBM, we prefer the term **"augmented intelligence".** This means that we aim to build systems that enhance and scale human expertise and skills rather than replacing them. We therefore focus on practical applications of discrete AI capabilities that assist people in performing well-defined tasks, by exploiting a wide range of AI-based services. We also use the term **"Cognitive Computing",** to mean a comprehensive set of capabilities based on technologies which include AI, but that go far beyond it. "Cognitive Computing"

Author: Francesca Rossi
Policy Department C: Citizens' Rights and Constitutional Affairs
European Parliament

comprises the fields of machine learning, reasoning and decision technologies, language, speech and vision recognition and processing technologies, human interface technologies, distributed and high-performance computing, and new computing architectures and devices. When purposefully integrated, these capabilities are designed to solve a wide range of practical problems, boost productivity, and foster new discoveries across many industries.

2. AI IN OUR LIVES

There are many examples of the presence of AI in our current life, that we don’t even know of. Whenever we buy something with a credit card, an AI algorithm approves that transaction (or not). When we use the GPS in our car, the algorithm that finds the best way to go from where we are to where we need to go is called the A* algorithm and it is an essential tool for AI, present in every AI teaching book. Spam filters are based on AI. Recommender systems such as that of Amazon are AI. The Google translate service, which nowadays is able to translate from and to more than 70 languages, is based on statistical machine learning, which is part of AI. Even web searches, such as those that we ask of Google or Baidu or other search engines, rely on AI to give us the web pages that are most relevant to our query. The face recognition capability of any of our cameras, shown usually with a green rectangle around each face we want to take a picture of, is AI. Siri, the IPhone app that understands us when we speak and responds (usually) in a useful way, is based on AI algorithms for speech understanding.

And of course there is the whole branch of robotics, which is more easily associated with AI because of the iconic image of humanoid robots that make it seem that humans have been reproduced artificially. Of course not all of them are intelligent in the way we would say a human is intelligent, but they are usually very good at doing what they are supposed to in their environment, from the Roomba robot that cleans the floors of our houses, to the Baxter robot that can work together with humans in production chains, passing through the Kiva warehouse robots that can take care of the tasks of an entire warehouse and the companion robots like Nao, Pepper, Aibo, and Giraff, who can entertain us, talk to us, and help elderly people to stay connected to their friends, relatives, and doctors.

The realm of possible uses of AI techniques is enormously vast, and this is one of the reasons why many companies have been heavily investing in AI in recent years. Google is building self-driving cars and has acquired more than 10 robotics companies, Facebook had opened a whole new research facility focused only on AI research, Apple has developed Siri, Microsoft has built Cortana, a similar personalized assistant, Google has acquired DeepMind, a UK company whose long-term aim is to build general AI and has already shown great potential in winning at the game of Go the current world champion, and IBM is investing a huge amount of resources in applying its Watson cognitive computing system to the medical domain, to finance, and to personalized education, just to name a few. This expansion of AI-based systems and services is reaching all corners of the globe. In Europe, IBM is establishing new centres in Munich and Milan focused on the application of cognitive computing capabilities to the Internet of Things and healthcare, respectively.

Self-driving cars are all about AI: they need to be able to see what happens in the street (signals, lanes, other cars, pedestrians, traffic lights), they need to able to predict what other cars and pedestrian will do, and they need to be able to cope with unforeseen situations. Since most car accidents are due to human fault, it is estimated that the adoption of self-driving cars will save about half of the lives that are usually lost in car accidents, which totals around 40,000 each year in the US alone. Some of us may be reluctant to hand over the wheel to an AI system, but very soon we may wonder why we did not do it sooner!

Watson is an IBM cognitive computing system that won against the best human champions at the Jeopardy! game in 2011. To do that, IBM Watson had to understand spoken language, make sense of massive amount of text, respond correctly to questions in many categories, as well as assess its own confidence in responding to such questions. The kind of question/answering capabilities that would be very useful, for example, in assisting a doctor when trying to come to the correct diagnosis for a patient and to propose the best therapy. Watson puts together many AI results and algorithms, from text and speech understanding, to reasoning with uncertainty, to optimization.

IBM is not new to tackling daunting challenges and successfully addressing them. In 1997, the Deep Blue computer program won against the world chess champion Garry Kasparov. This was very iconic, since chess, as I said above, is one of those activities that we believe requires a significant amount of intelligence in a human being. Deep Blue showed that computers could do better than the best humans when it comes to certain tasks.

3. AI AND COMPUTING POWER

We have to be careful in labelling all this promising progress as truly "intelligent." Humans need intelligence and good intuition in playing chess because our brain does not have enough computing power to make sense of a lot of data. In chess, for example, it is not possible for our brain to evaluate all possible sequences of moves of us and our opponent in a very short time. If we could do that, it would be obvious to us what the best move is. Contrarily to our brain, computers can rely on a computing power that, according to Moore's law, doubles every about 18 months. Gordon Moore, co-founder of Intel, in 1965 noticed that this was the trend in putting transistors into a single chip, and to the amazement of many, this law has been followed since then in our computers.

This law means that computer processing power doubles every 18, or, seen from another point of view, every 18 months we can have a much cheaper computer with the same speed as the old one. For example, it has been calculated that an IPhone in 1991 would have cost about $3,6 million. And this is only for its processor, its memory, and its connectivity. Today's smart phones are more powerful than NASA computers that in 1969 sent a man to the moon. This is how much computing power has increased over the years.

4. AI AND DATA

AI is not all about computing power. Intelligent machines can also rely on huge amounts of data, to be used to learn how to make better and better decisions. This data comes from all of us. Over the years, Facebook users have uploaded more than 250 billion pictures, and every day they upload about 350 million more. Every second, we submit 40,000 Google search queries, which means 3.5 billion per day and 1.2 trillion per year. As of today, there are 2 billion people connected to internet, which is estimated to get to 5 billion by 2020. And by that time, also 50 billion "things" will be connected through the web: from appliances to traffic lights, from cars to watches.

5. MACHINES VS HUMANS

No matter how much data and computing power is available to machines, there are tasks that are still difficult for machines to perform, but that remain very easy for humans. Machines and humans are very complementary. A typical example is understanding what is depicted in an image.

How do we know that an image contains a cat? Because during our life we have seen many examples of cats and non-cats, and at some point we got a very good idea of how a cat should look like, so much that we don't have problems recognizing one even if we have never seen it before, and even if it is in a strange position.

Machines need humans to provide them with many examples. A lot of progress has been made in this, but we are still working hard to improve their accuracy in labelling pictures or other perception capabilities

Other tasks that are very easy for humans are physical and manipulation tasks such as walking, running, picking up an object no matter its shape and location. Robots can do this only in restricted environments. But they are still not able to have the general physical and manipulation capabilities even of a 6 year old.

6. AI ETHICS AND TRUST

The ability of AI systems to transform vast amounts of complex, ambiguous information into insight has the potential to reveal long-held secrets and help solve some of the world's most enduring problems. AI systems can potentially be used to help discover insights to treat disease, predict the weather, and manage the global economy. It is an undeniably powerful tool. And like all powerful tools, great care must be taken in its development and deployment. However, to reap the societal benefits of AI systems, we will first need to trust it. The right level of trust will be earned through repeated experience, in the same way we learn to trust that an ATM will register a deposit, or that an automobile will stop when the brake is applied. Put simply, we trust things that behave as we expect them to.

Trust is built upon accountability. As such, the algorithms that underpin AI systems need to be as transparent, or at least interpretable, as possible. In other words, they need to be able to explain their behaviour in terms that humans can understand — from how they interpreted their input to why they recommended a particular output. To do this, we recommend all AI systems should include explanation-based collateral systems.

But trust will also require a system of best practices that can help guide the safe and ethical management of AI systems including alignment with social norms and values; algorithmic responsibility; compliance with existing legislation and policy; assurance of the integrity of the data, algorithms and systems; and protection of privacy and personal information.

One of the primary reasons for including algorithmic accountability in any AI system is to manage the potential for bias in the decision-making process. This is an important and valid concern among those familiar with AI. Bias can be introduced both in the data sets that are used to train an AI system, and by the algorithms that process that data. At IBM, we believe that the biases of AI systems can not only be managed, but also that AI systems themselves can help eliminate many of the biases that already exist in human decision-making models today.

AI systems should function according to values that are aligned to those of humans, so that they are accepted by our societies and by the environment in which they are intended to function. This is essential not just in autonomous systems, but also in systems based on human-machine collaboration, since value misalignment could preclude or impede effective teamwork. It is not yet clear what values machines should use, and how to embed these values into them. Several ethical theories, defined for humans, are being considered (deontic, consequentialist, virtue, etc.) as well as the implications of their use within a machine, in order to find the best way to define and adapt values from humans to machines.

In industries like healthcare and finance, the relevant professional ethical principles are explicitly encoded and practiced by professionals in the field already. In AI systems designed to help professionals in these domains, these best practices and principles could form the core of the ethics module for such systems. Ethics modules, however, should be constantly adapted to reflect humans' best practices in their everyday profession.

We envision a future in which every AI system will need to have its own ethics module to allow for a fruitful interaction and collaboration with humans in the environments in which it is used. This could be achieved by developing an ethics API that can be adapted to specific professions and real-life scenarios. It would provide the main principles and values the AI systems should base its behaviour on, as well as the capability to dynamically adapt them over time to tune them to the real situations that are encountered in that profession or environment. Such a rigorous approach could offer sufficient value alignment without compromising the full problem-solving potential of artificial intelligence.

7. IBM AND AI

IBM has been researching, developing and investing in AI technology for more than 50 years. In 1997, IBM Deep Blue bested then world chess champion Garry Kasparov, showing that innovative AI algorithms and computational power can play a complex game at super-human levels. In 2011, IBM Watson won at Jeopardy! against the best human players, showing that AI can also perform very well in natural language understanding and reasoning with uncertainty.

These are just the tip of the iceberg compared to what IBM has been achieving over the years in the field of AI. We have been transforming the original Watson program into a fully fledged platform and we have exploited it to successfully apply AI to many industrial sectors, including healthcare, finance, commerce, education, security, and the Internet of Things. The whole company is deeply committed to AI, since we believe strongly in its potential to benefit society while transforming our personal and professional lives.

As mentioned above IBM prefer the term “augmented intelligence” and therefore focuses on practical applications of AI capabilities that assist people in performing well-defined tasks, by exploiting a wide range of AI-based services. With this aim in mind, IBM researchers, in tight collaboration with several universities, produce continuous innovations in area such as machine learning, knowledge modelling, reasoning and decision technologies, human interface, automated perception, data assurance, and computing infrastructures. Most of these research efforts cannot be achieved by AI researchers alone. Collaboration with experts in multiple disciplines — such as psychology, philosophy, sociology, art, regulation, and law — is crucial.

We believe that new companies, new jobs, and entirely new markets will be built on the shoulders of this powerful technology. Moreover, AI systems will improve access to critical services for underserved populations. Overall, we anticipate widespread improvements in the quality of life.

8. IBM AND AI ETHICS

In order to be fully accepted into society, AI systems need to have significant social capabilities, because their presence in our lives has a profound impact on our emotions and on our decision making capabilities. Also, AI systems need to understand how to learn and comply with specific behavioural principles for aligning with human values. To take full advantage of the potential societal benefits of AI, we will need to trust AI, whether we speak of autonomous systems or, as is the focus of IBM, of human/machines partnerships. Trust will be earned over time and via natural interaction modalities. Trust will also require a system of best practices that can guide the safe and ethical development and management of AI, a carefully thought alignment with social norms and values, algorithmic accountability, compliance with existing legislation and policy, and protection of privacy and personal information.

IBM is in the process of developing this system internally, with our collaborators, and also with our competitors. More precisely, IBM is engaged in several efforts – both internally and externally – to advance our understanding and effecting the ethical development of artificial intelligence. They include:

- The establishment of an internal IBM Cognitive Ethics Board, to discuss, advise and guide the ethical development and deployment of AI systems.
- A company-wide educational curriculum on the ethical development of cognitive technologies.

- The creation of the IBM Cognitive Ethics and Society research program, a multi-disciplinary research program for the ongoing exploration of responsible development of AI systems aligned with our personal and professional values.
- Participation in cross-industry, government and scientific initiatives and events around AI and ethics, such as the recently launched Partnership on AI, the White House Office of Science and Technology Policy AI workshops, the International Joint Conference on Artificial Intelligence, and the conference of the Association for the Advancement of Artificial Intelligence.
- Regular, ongoing IBM-hosted engagements with a robust ecosystem of academics, researchers, policymakers, NGOs and business leaders on the ethical implications of AI.

9. IBM AND EUROPE

IBM is an international company with a strong history and presence in Europe:
The IBM Zurich research lab is supported by a multicultural and interdisciplinary team of a few hundred people from about 45 nationalities who work in diverse areas such as chip technologies, nanotechnology, fibre optics, supercomputing, data storage, security and privacy, risk and compliance, business optimization and transformation, and server systems. The Zurich lab is involved in many joint projects with universities throughout Europe, in research programs established by the European Union and the Swiss government, and in cooperation agreements with research institutes of industrial partners.

The recently opened IBM Watson IoT Headquarters in Munich is applying Watson to the Internet of Things and helping many companies (from automotive, insurance, electronics, banks, and industrial sectors) to transform their business by extending the power of AI to the billions of connected devices, sensors and systems that comprise the IoT.

The recently announced IBM Watson Health European Centre of Excellence, that will be placed within the Human Technopole Lab in Milan, is supporting the government of Italy's initiative to establish an international hub for the advancement of genomics, big data, aging, and nutrition. The Centre is expected to provide access to resources and technology designed to help accelerate research into new treatment options, promote personalized medicine, and encourage discoveries aimed at improving overall public health management while advancing sustainable health systems.

10. AI AND POLICIES

AI technology is changing so rapidly, and has so many applications to the real world, that it is difficult for any government or regulatory agency to keep up with them and to meaningfully and timely guide the deployment of AI systems. However, some issues like data privacy and ownership have been considered in the EU, as well as algorithm transparency and accountability.

An example is the recently released General Data Protection Regulation, that will take effect as law across the EU in 2018 and will restrict automated individual decision-making (that is, algorithms that make decisions based on user-level predictors) which "significantly affect" users. The law will effectively create a so-called "right to explanation," whereby a user can ask for an explanation of an algorithmic decision that was made about them. Another example is the very recently released USA federal policy on automated vehicles, that is already in effect.

The main point of all these policies is to make sure that society can take full advantage of the capabilities of AI systems while minimizing the possible undesired consequences on people. Safety is very important, as well as fairness, inclusiveness, and equality. These and other properties should be assured of AI systems, or at least we should be able to assess the limits of an intelligent machine, so to not overtrust it. It is therefore very important the policies and regulations help society in using AI for the best of all.

Ethical issues, including safety constraints, are essential in this respect, since an AI system that behaves according to our ethical principles and moral values would allow humans to interact with it in a safe and meaningful way.

It is clear that a lack of regulations would open the way to unsafe developments. However, also excessive regulations would have a cost to society, since they would not allow us to take advantage of all the potential benefits that AI can bring, such as saving lives, curing diseases, and solving planetary problems.

IBM is eager to work with governments, media, other companies, regulatory agencies, and industry sectors in a meaningful discussion on ethical issues of AI, with the aim of clearly identifying the potential and limits of AI, and carefully understanding how to harness it for the best of all.

DISCLAIMER

The content of this document is the sole responsibility of the author and any opinions expressed therein do not necessarily represent the official position of the European Parliament. It is addressed to Members and staff of the EP for their parliamentary work. Reproduction and translation for non-commercial purposes are authorised, provided the source is acknowledged and the European Parliament is given prior notice and sent a copy.

This document is available at: www.europarl.europa.eu/studies
Contact: poldep-citizens@ep.europa.eu
Manuscript completed in October 2016
© European Union

PE 571.380

CATALOGUE: QA-06-16-133-EN-C (paper)
CATALOGUE: QA-06-16-133-EN-N (pdf)
ISBN: 978-92-846-0122-6 (paper)
ISBN: 978-92-846-0121-9 (pdf)
doi: 10.2861/730206 (paper)
doi: 10.2861/31941 (pdf)