BRIEFING
Requested by the IMCO committee

European Parliament

State of the art and future of artificial intelligence

KEY FINDINGS

Artificial Intelligence is a booming field of scientific discovery and practical deployments. Once a mostly academic area of study, twenty-first century AI enables a spectrum of mainstream technologies that are having a substantial impact on everyday lives. In many cases, already now, AI accompanies the users in our everyday errands and professional lives. In the future it will not only reshape business, public administration, health care, finances or education, but may also contribute to solving grand civilizational challenges such as climate change, hunger or inequality. The phase of AI massively transforming society, economy, and politics has already begun.

Introduction

The term "Artificial Intelligence" (AI) is notoriously hard to define. Frequently people use it to mean things that are hard for computers to do (like understanding natural language) as opposed to things that we know computers handle quite well (like accounting). What is artificial intelligence? One can start with a very simple definition: machines acting in ways that seem intelligent. This, however, is currently already insufficient. Thus, it is more suitable to say that artificial intelligence (sometimes called machine intelligence or designed intelligence - Jang et al. 1997, Heift 1998), is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and animals. One could also state that artificial intelligence is an umbrella term embracing computer (machine) vision, natural language processing, virtual assistants and bots, robotic process automation, machine learning (including most advanced techniques like deep learning), and cognitive processes in organizations (Simon 1996) (Rasmussen, Pejtersen, and Goodstein 1994) (Woods and Hollnagel 2005) (Ogiela and Ogiela 2012). Moreover, artificial intelligence is a **branch of science and as such it can be defined as a set of computational technologies that are inspired by the ways people use their nervous systems and bodies to sense, learn, reason, and take action**.
This definition will be the one we should treat as most suitable for the purposes of this briefing paper.

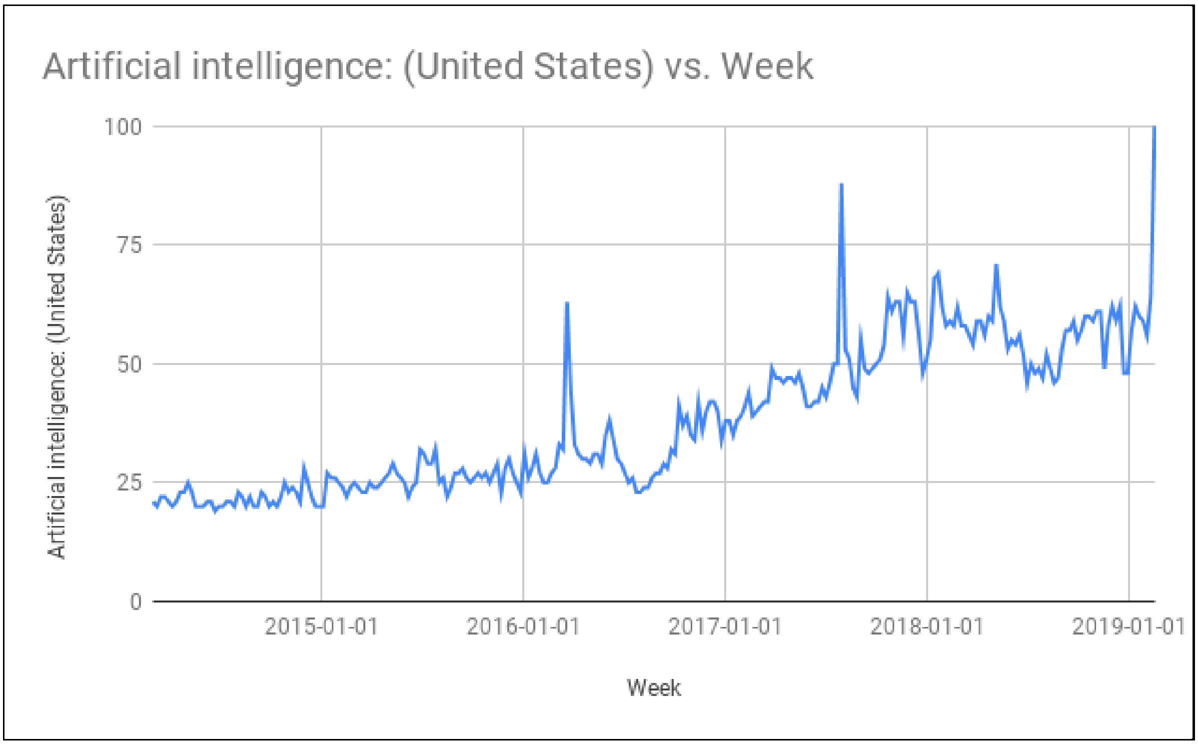
AI also needs to be viewed and defined on the temporal axe, as within this umbrella term, there are many "moving" objects. What was considered "artificial intelligence" back in time, currently does not necessarily count as AI.

Policy Department for Economic, Scientific and Quality of Life Policies
Directorate-General for Internal Policies
Author: Dr. Aleksandra Przegalinska
PE 631.051- February 2019

EN

A good example of such "moving object" is natural language processing, today perceived more as a subfield of computer science rather than artificial intelligence.

Figure 1: Growing Interest on Artificial Intelligence



Source: Own calculations (on the basis of Google Trends as of 26.02.2019)

Discussions regarding artificial intelligence date back to the 1940s and 1950s. Since the field's inception sixty years ago the rate of progress in AI has been patchy and unpredictable. Its fairly long history includes both periods of abundant funding and vital interest (for the sake of this paper we may dub them "AI summers") and periods of disappointment and general lack of investments - the so-called "AI winters" (Hendler 2008). Many of the inventions in AI that general public perceives as entirely new also date back to some decades ago. A good example here are artificial neural networks that have been conceptually developed first in the 40ies of the 20th century, later on in the 80ies and currently reemerged and are becoming most important paradigm within artificial intelligence per se.

If we take into account historical developments in AI, we can enumerate three main paradigms within the field: (1) **AI** - consisting, among others, in inductive Logic Programming, robotic process automation and expert systems (including fuzzy systems), (2) **statistical AI** - consisting in decision networks, probabilistic programming (including Bayesian Program Synthesis), computer vision (activities and image recognition as well as machine vision), natural language processing and machine learning, and, last but not least, (3) **subsymbolic AI** - consisting in distributed artificial intelligence (including agent-based modeling, swarm intelligence and multi-agent systems), ambient computing, affective computing, embodied intelligence and various autonomous systems (Knight 1989), (Jang et al. 1997), (Munakata 2001), (Wu et al. 2007).

Recent developments in Artificial Intelligence and applications in business

Nowadays, AI-based systems can be found in a variety of applications from smartphones to CRM-systems and stock exchanges. A wide range of AI's business applications, stemming from the symbolic, subsymbolic and statistical paradigms, embrace, among others, financial risk assessment, pricing optimization, customer targeting and personalization of service, medical diagnosis, recommender systems and virtual assistants. As far as the most widely applied discipline within AI, machine learning, is concerned, we can distinguish 3 pillars: supervised learning, unsupervised learning and reinforcement learning. Supervised learning, consist of such methods like **regression**, frequently used in population growth prediction or estimating life expectancy and **classification**, used, among others, in identity fraud detection. Unsupervised learning consists of dimensionality reduction that allows for visualization of big data, and clustering, frequently used by business in customer segmentation. Reinforcement learning, is used in robot navigation, robot skill acquisition and game AI (AI playing games, including chess, Go or video games).

Executives in companies around the world are increasingly looking to artificial intelligence to create new sources of business value. This is especially true for pioneering adopters of AI - those that have invested in AI initiatives and seen impressive results. This small group of companies is doubling down on AI investments, building competencies, and working to take AI to scale. The opportunities and challenges these AI Pioneers face are the focus of the 2018 MIT Sloan Management Review and The Boston Consulting Group (BCG) as well as previous similar research reports (Ransbotham et al. 2017) (Ransbotham et al., n.d.). In the survey, the authors and analysts classified the organizations into four groups based on their responses to questions about their levels of AI adoption and AI perception. (1) Pioneers are enterprises that have high understanding of AI concepts and methods, as well as significant levels of AI deployment; (2) investigators understand AI but have limited adoption; (3) experimenters have adopted AI but do not fully grasp the context; and (4) "passives" have little understanding and adoption of AI.

In order to be able to address the rapid development of this technology, make most of its potential and avoid risks, we should prepare in advance. One very prominent example of such ambiguity is related to AI-based autonomous vehicles (self-driving cars). Such vehicles combine a variety of sensors to perceive their surroundings, such as Lidar, sonar, radar, GPS, odometry and inertial measurement units. Beyond that, they are equipped with computer (machine vision) that currently is one of the most prominent AI technologies. Systems based on artificial intelligence used in autonomous transportation can substantially reduce the number of accidents and, as a result, reduce the number of fatalities. AI systems can contribute to achieving better and more accurate results in health diagnoses. Additionally, the implementation of simple robotics in companies may lead to noticeable savings, not to mention data-driven economy, where the use of AI solutions plays a significant role. This is a very promising area for European startups, especially in the financial, judicial and health sectors. In fact, the challenge of making industrial data accessible for re - use in Europe is as much imminent as it is inevitable. AI is all about data-based innovation, after all. We may also be looking at supporting disruptive innovation. AI could strongly support the reform of public administration decision-making tools, e.g. when preparing regulatory impact assessment and this could create new demand.

Future developments in artificial intelligence

In terms of future developments of artificial intelligence there are a quite few trends worth mentioning. Apart from the already expected developments in chatbots and virtual agents (including the recent initiative by DeepMind to build a Theory of Other Mind Network able to understand changing mind states of interlocutors[1]) and natural language usages in business intelligence, one can also mention machine learning-empowered cognitive analytics, where machines learn from experience and gradually build context awareness (ambient intelligence). Another relevant trend is related to parallel information processing, aided through chips designed for AI applications.

With the advances in the Internet of Things, sensors and wearable technologies, smart object and environments are another important trend that AI can significantly enhance through collected big data analytics. Deep learning approaches allow processing of raw data including textual data, speech and images. Already now some experts ask, however, what is behind deep-learning and its current limitations and whether we can expect the arrival of common-sense understanding AI[2]. Finally, intelligent automation and automated machine learning are two emerging combinations of automation with artificial intelligence that allow various knowledge workers, to process, understand and use volumes of information to enhance their professional activities.

Chatbots and social robots an example of successful subfields of AI

As mentioned above, chatbots are on the frontline of innovation and future developments of artificial intelligence. They are increasingly frequently used in business to facilitate various processes, particularly those related to customer service and personalization. They also harvest data about Internet users and spread information online.

Chatbots are computer programs built to engage with a human interlocutor, for instance through received messages. Chatbots can be programmed in diverse ways. Rule based chatbots respond the same way each time. Chatbots based on machine learning (in particular deep learning) are able to respond differently to messages containing certain keywords and even adapt their responses to fit particular context. Chatbots leverage chat mediums social messaging services, SMS messages, or website chat windows. They can be used for finding customer service assistant, making reservations, paying bills, buying items online, or interacting with brands and organizations. They come in all forms, but are generally classified as humanoid AI. Humanoid bots and robots perform certain activities as a substitute for humans, whose function is often to imitate human behavior.

A typical humanoid disposes of machine learning, visual data processing and a facial recognition system. Similarly to chatbots, social robots that are also considered humanoid AI, possess the same features but without physical resemblance to a human. Chatbots and social robots may imitate human gestures and facial expressions. One of the most famous examples of humanoid robots is Sophia, created by Hanson Robotics, that was granted citizenship of Saudi Arabia. Sophia uses Alphabet's voice recognition technology and is designed to become smarter with the passage of time. Interestingly enough, on the level of natural language usage Sophia is a chatbot. It is actually conceptually similar to the ELIZA bot program which was one of the first attempts to simulate human conversation.

In 1966, designed by Joseph Weizenbaum, ELIZA, the first bot capable of talking to people, conducted several „therapeutic" conversations with patients, acting as a Rogerian psychologist. ELIZA was a bot that immensely inspired the so-called the Turing test - an experiment that was conceived as a way to determine

[1] See also: https://deepmind.com/research/publications/machine-theory-mind/

[2] See also: https://www.wired.com/story/how-to-teach-artificial-intelligence-common-sense/

the machine's ability to use natural language and indirectly to prove its ability to think in a way similar to human.

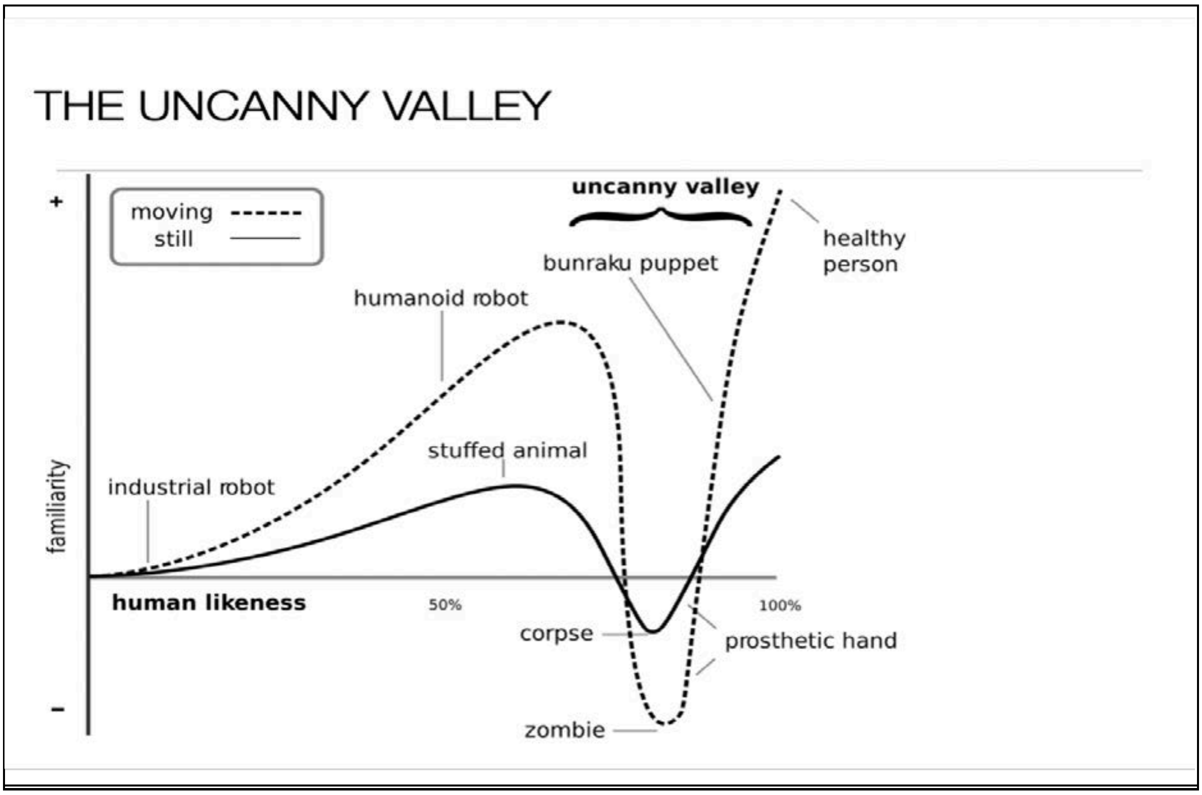
Until now bots are competing in annual Turing Test-like challenges. The newest deep-learning based bots, such as Duplex by Google are so skillful while interacting with humans that they are barely recognizable as machines, at least in shorter conversations.

Chatbots are a very well timed middle ground between fairly simple automation, artificial intelligence, and the emerging interfaces of communication. Already now they prove to be extremely effective in conducting business online and ushering in some of the other future technological advancements.

“Gartner predicts that by 2020 people will have more conversations with chatbots than their spouse,” said Christi Olson, chatbot expert and head evangelist at Bing. She also added that “the chatbots of the future don’t just respond to questions. They talk. They think. They draw insights from knowledge graphs. They forge emotional relationships with customers.”[3]

The work of the team from MIT Center for Collective Intelligence, SWPS University of Psychology and Kozminski University is circling around anthropomorphic bots and robots, often referred to as humanoids. In the current study, the focal point of research is a very known and researched Uncanny Valley effect, described in human-machine interaction literature as a feeling of eeriness accompanying human interaction with humanoids. By integrating the methodology of psychological questionnaires with electrophysiological measurements, the research team managed to confirm the significant effect of Uncanny Valley (Fig. 1) in those bots that present more anthropomorphic features.

Figure 2: Effect of the Uncanny Valley



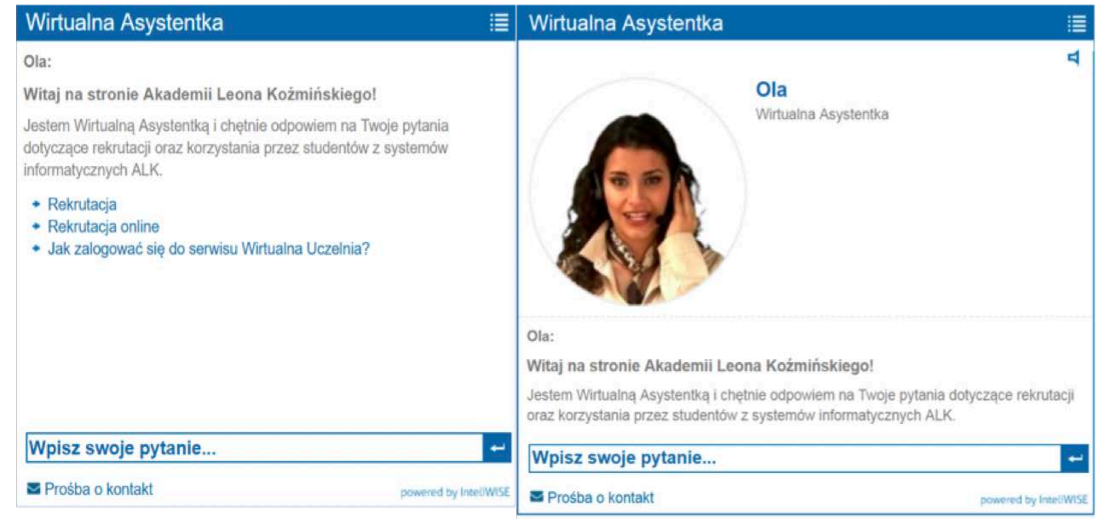
Source: https://www.sciencedirect.com/science/article/pii/S0167739X17312268

[3] See also: https://www.searchenginejournal.com/future-of-chatbots/278595/

The experiment itself consisted of two parts: measurement of psychophysiological reactions of chatbot users and a detailed questionnaire that focused on assessing interactions and willingness to collaborate with a bot.

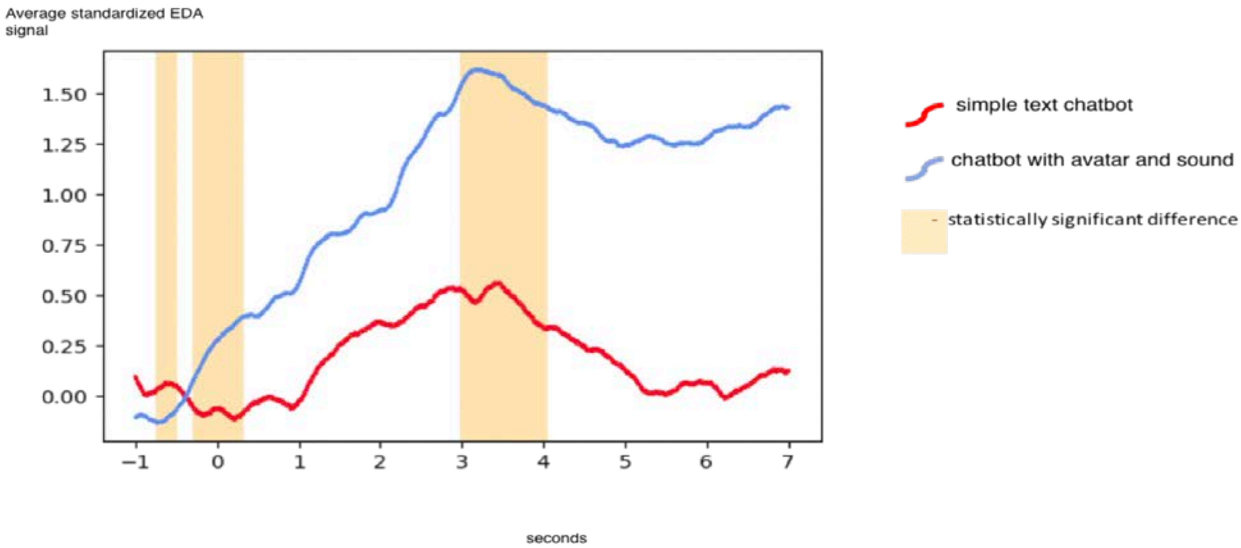
In the first quantitative stage, participants interacted with a chatbot, either with a simple text chatbot (control group) or an avatar reading its responses in addition to only presenting them on the screen (experimental group). The results have shown that participants were experiencing lesser uncanny effects and less negative affect in cooperation with a simpler text chatbot than with the more complex, animated avatar chatbot. Simple Chatbot have also induced less intense psychophysiological reactions. Thus, understanding the user's side may be crucial for designing better chatbots in the future and, as many other studies researching new forms of human-machine interaction, can contribute to advancing the field of human–computer interaction.

Figure 3: Humanoid chatbot in text and avatar form (Polish version of the chatbot welcoming students to Kozminski University's Dean's Office and presenting frequently asked questions. The section below the chatbot is an editor where students can post their queries).



Source: https://www.sciencedirect.com/science/article/pii/S0167739X17312268

Figure 4: Average standardized EDA Response



Source: https://www.sciencedirect.com/science/article/pii/S0167739X17312268

In the current research the proposed novel methodology of tracking human-chatbot interactions and measuring chatbot performance takes into consideration ethical concerns. It links neuro-scientific methods, text mining and deep learning with the issue of trust allocated in chatbots and their overall transparency. This technology consists in analysing the content of messages produced in human-chatbot interactions, using the Condor Tribefinder system developed by Prof. Peter Gloor's team for text-mining (De Oliveira and Gloor 2018; Gloor et al., n.d.) and based on a machine learning classification engine.

Moreover, it is argued that trust is the focal point of successful human-chatbot interaction and that trust as a category is redefined with the advent of deep-learning supported chatbots.

This analysis allowed for clear delineation of values shared by the chatbot that were extracted from conversations it had with humans. This discovery leads to a conclusion that also chatbots can transmit and share certain ideologies and values and represent certain interests. More importantly, another part of the research showed that chatbots were vulnerable to quick ideological changes that made them declare juxtaposed views in one short conversation - behavior highly unlikely for humans that usually present a high degree of ideological integrity, at least in the short term.

A major relevant question arising is related to trust allocated in chatbots and other natural language processing entities that handle human knowledge. Would we expect such entities to be predictable and present integrity or rather be flexible and follow the users no matter what kind of ideas they present? Will that matter for future generations, as Stefania Druga and Richard Williams rightfully ask in their cutting-edge research on children's interactions with virtual assistants and social robots[4] called „Hey Google is it okay if I eat you?"

What is certain is that chatbots are the perfect example of the implementation of state-of-the-art consumer-oriented artificial intelligence that does not only simulate human behavior – based on formal models – but adapts to it. As such, they are a fascinating subject for the research of patterns in human and non-human interaction along with issues related to assigning social roles to others, finding patterns of (un)successful interactions, and establishing social relationships and bonds.

Main Current Challenges in Artificial Intelligence

As much as we can appreciate growing number of pioneers and investigators over the past few years, impacts of the AI development are not limited to business and industrial development. They also have a significant legal, social and ethical dimension. While on the one hand, AI offers a huge potential, we are at the same time facing the challenge of re-skilling, legal and ethical and legal dilemmas regarding implementation of algorithmic decision making (Corbett-Davies et al. 2017; Diakopoulos 2016), and issues related to overall system explainability, particularly visible in deep learning. Among some of the most crucial AI-related challenges one should mention:

Accessibility and integrity: AI systems will be increasingly used in significant, if not critical contexts. Various adversaries can threaten the integrity or availability of machine learning-based in different ways i.e by polluting training datasets with fake data or attacking the ML algorithm.

Privacy and safety: privacy concerns are a relevant problem for the functioning of AI-based systems. Improper or non-existent disclosure control can be the cause for data privacy issues. Such issues may arise in response to information from a wide range of sources, such as financial transactions, healthcare records,

[4] See also: https://www.media.mit.edu/publications/hey-google-is-it-ok-if-i-eat-you-initial-explorations-in-child-agent-interaction/

or revealing biological traits. All the aforementioned types of data are frequently processed by AI-based systems.

Tackling algorithmic bias: such bias can be found across platforms, including but not limited to search engine results and social media platforms. It occurs when a computer system reflects the implicit values of the humans who are involved in coding, collecting, selecting, or using data to train the algorithm. The bias can have impacts ranging from inadvertent privacy violations to reinforcing social biases of race, ethnicity, sexuality or gender. The study of algorithmic bias is most concerned with algorithms that reflect "systematic and unfair" discrimination. Current solutions focus on the so-called “protected features”.

Explainability and transparency refers to the type of AI whose actions can be easily understood by humans. It contrasts with the concept of the “black box” in deep learning, ("interpretability" of the workings of complex algorithms, where even their designers cannot explain why the AI arrived at a specific decision). Explainable AI can also be used to implement a social right to explanation. It is worth noting that some experts claim that transparency comes at a price and that there are often tradeoffs between how advanced AI is and its general transparency. Another consideration is *info-besity* (overload of information), thus, *full transparency* may not be always possible or even required. The amount of information presented should vary based on the stakeholder interacting with the intelligent system.

Current state of technological development does not clarify what will be the next stage of affairs and what sort of use will we make of those technologies that are either replacing people or opening up a new, radically deeper level of machine - human interaction and interdependency. Humans are no longer sure if they are indeed following the path of humanizing technology, or rather moving towards adapting humans to technologies. Recent scandals with major data leaks at Facebook, Grindr, accidents involving autonomous vehicles, threats of autonomous weapons and other problems do not make understanding of this complex process any easier. Three most important steps for the Artificial Intelligence/ data scientific community, as well as the broader audience consist in:

- proposing **AI ethics guidelines** , covering issues such as fairness, safety, transparency, the future of work, democracy, privacy and personal data protection, dignity, as well as non-discrimination.
- incentivizing the creation of European centers of excellence that could contribute to **technical solutions** of the problems of explainability and fairness in AI.
- examining possible funding schemes in order to increase competitiveness of AI research and development in Europe.

Many of these issues are already discussed in-depth in the European Commission´s Coordinated plan on Artificial Intelligence (AI) Communication[5] .

[5] See also: http://europa.eu/rapid/press-release_IP-18-6689_en.htm

REFERENCES

- Ciechanowski, Leon, Aleksandra Przegalinska, Mikolaj Magnuski, and Peter Gloor. 2018. "In the Shades of the Uncanny Valley: An Experimental Study of Human–chatbot Interaction." *Future Generations Computer Systems: FGCS*, February.
  https://doi.org/10.1016/j.future.2018.01.055
  Corbett-Davies, Sam, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. 2017. "Algorithmic Decision Making and the Cost of Fairness." In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 797–806. KDD '17. New York, NY, USA: ACM.
- De Oliveira, Joao Marcos, and Peter A. Gloor. 2018. "GalaxyScope: Finding the 'Truth of Tribes' on Social Media." In *Collaborative Innovation Networks: Building Adaptive and Resilient Organizations*, edited by Francesca Grippa, João Leitão, Julia Gluesing, Ken Riopelle, and Peter Gloor, 153–64. Cham: Springer International Publishing.
- Diakopoulos, Nicholas. 2016. "Accountability in Algorithmic Decision Making." *Communications of the ACM* 59 (2): 56–62.
- Gloor, Peter, Andrea Fronzetti Colladon, Joao Marcos de Oliveira, and Paola Rovelli. n.d. "Identifying Tribes on Twitter through Shared Context." http://suzhou18.coinsconference.org/doc/COINs18_paper_2.pdf.
- Heift, Gertrud Doris M. 1998. Designed Intelligence: A Language Teacher Model. Simon Fraser University.
- Hendler, J. 2008. "Avoiding Another AI Winter." *IEEE Intelligent Systems* 23 (2): 2–4. Jang, S. R. J., Sun, C. T, Mizutani, and E. 1997. "Neuro-Fuzzy and Soft Computing; a Computational Approach to Learning and Machine Intelligence." http://cumincad.scix.net/cgi-bin/works/Show?d036.
- Knight, Kevin. 1989. *A Gentle Introduction to Subsymbolic Computation: Connectionism for the AI Researcher*. Carnegie Mellon University, Computer Science Department. Munakata, Toshinori. 2001. *Fundamentals of the New Artificial Intelligence: Beyond Traditional Paradigms*. Springer Science & Business Media.
- Ogiela, Lidia, and Marek R. Ogiela. 2012. "Cognitive Systems and Artificial Brains." In *Advances in Cognitive Information Systems*, edited by Lidia Ogiela and Marek R.
- Ogiela, 99–106. Berlin, Heidelberg: Springer Berlin Heidelberg.
  Ransbotham, Sam, Philipp Gerbert, Martin Reeves, David Kiron, and Michael Spira. n.d. "Artificial Intelligence in Business Gets Real." https://sloanreview.mit.edu/projects/artificial-intelligence-in-business-gets-real/.
- Ransbotham, Sam, David Kiron, Philipp Gerbert, and Martin Reeves. 2017. "Reshaping Business With Artificial Intelligence: Closing the Gap Between Ambition and Action." *MIT Sloan Management Review* 59 (1). http://search.proquest.com/openview/83d554491afeb2435c6c2e386821c60c/1?pq-origsite=gscholar&cbl=26142.
- Rasmussen, Jens, Annelise Mark Pejtersen, and Len P. Goodstein. 1994. "Cognitive Systems Engineering."
- Simon, Herbert A. 1996. The Sciences of the Artificial. MIT Press.
- Woods, David D., and Erik Hollnagel. 2005. Joint Cognitive Systems: Foundations of Cognitive Systems Engineering. CRC Press.
- Wu, Q., D. A. Bell, G. Prasad, and T. M. McGinnity. 2007. "A Distribution-Index-Based Discretizer for Decision-Making with Symbolic AI Approaches." *IEEE Transactions on Knowledge and Data Engineering* 19 (1): 17–28.
- Woods, David D., and Erik Hollnagel. 2005. Joint Cognitive Systems: Foundations of Cognitive Systems Engineering. CRC Press.

Disclaimer and copyright. The opinions expressed in this document are the sole responsibility of the authors and do not necessarily represent the official position of the European Parliament. 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. © European Union

Manuscript completed: January 2019; Date of publication: February, 2019
Administrator responsible: Mariusz MACIEJEWSKI; Editorial assistant: Irene VERNACOTOLA
Contact: Poldep-Economy-Science@ep.europa.eu
This document is available on the internet at: www.europarl.europa.eu/supporting-analyses

IP/A/IMCO/2018-138
Print ISBN 978-92-846-4606-7 | doi: 10.2861/25781 | QA-03-19-103-EN-C
PDF ISBN 978-92-846-4605-0 | doi: 10.2861/613707 | QA-03-19-103-EN-N