LUCA AI Powered Decisions

Six challenges for Artificial Intelligence's sustainability and what we should do about it

Dr. V. Richard Benjamins, Telefónica

luca-d3.com

Telefónica DATA UNIT

Índex

Whitepaper

1. Introduction

Artificial Intelligence (AI) is on the rise. There are so many innovative applications of AI that everybody is speaking about it. It improves the diagnosis and treatment of cancer, improves customer experience, creates new business, improves education, predicts how contagious diseases propagate and optimizes the management of humanitarian catastrophes, to name just a few.

Artificial Intelligence already exists since the mid-fifties when John McCarthy first coined the term. Marvin Minsky's definition of AI was: "The science of making machines do things that would require intelligence if done by humans." AI has had its ups and downs previously (e.g. AI Winters). In a white paper published earlier[1], we explained basic concepts of AI to provide general readers with some background information to better understand the many articles about AI appearing in the press and Internet. AI's current popularity is mostly due to the enormous progress in one of AI's subfields: Machine Learning (ML). There are three main reasons for this progress:

- The abundance of data. ML analyses data, and today there is so much more data available than decades ago.
- Increase in computational power. Moore's Law is still valid and machines can process orders of magnitude faster and more than before.
- Deep Learning. An extended version of Neural Networks that, thanks to the two previous points, has increased enormously the performance of all kinds of classification and prediction tasks.

However, with all those opportunities also comes great responsibility to ensure good and fair practice of AI. We identified six societal and ethical challenges for Artificial Intelligence that should be dealt with before AI is massively applied. Not treating those issues will lead to uncertainty and unforeseen consequences with potentially large negative societal impact. It is therefore no surprise that many governments currently have set up national initiatives to discuss many of those issues (e.g. in the UK[2] and France[3]).

2. Six societal and ethical challenges for Data and AI

The challenges that AI, and therefore Data as well, face, include non-desired side effects, liability questions, yet unknown consequences, the relation human-robots, increasing concentration of power and wealth, and intentional bad uses. Below, we explain each of them.

1. Non-desired side effects.
    - While Machine Learning is able to solve complex tasks with high performance, it might use information that from a society or humanity perspective is undesired. For example, deciding whether to provide a loan to people based on race or religion is not something our societies accept. While it is possible to remove those "undesired" attributes from data sets, there are other less obvious attributes that highly correlate with those "undesired" attributes whose removal is less

straightforward. Machine Learning is objective and finds whatever relation there is in the data regardless of specific norms and values.

- A related issue is so-called bias of data sets. Machine Learning bases its conclusions on data. However, the data itself might be biased by not being representative for the group of people to which the results are applied. For instance, finding trends on school performance using mostly white schools will not provide insights applicable to all schools. Research has shown that ML takes over any bias from humans.[4,5]

- Apart from bias in the training data, bias can also come from the algorithm. A Machine Learning algorithm tries to be as accurate as possible when fitting the model to the training data. Accuracy can be defined in terms of so-called "false positives" and "false negatives", often through a so-called confusion matrix[6]. But the definition of this "accuracy" measure, whether it tries to optimize only false positives or only false negatives, or both, has an important impact on the outcome of the algorithm, and therefore on the groups of people affected by the AI program. In safety-critical domains such as health, justice, and transport defining "accuracy" is not a technical decision, but a domain or even a political decision.

- Deep learning algorithms can be highly successful but have a hard time to explain why they have come to a decision. For some applications, the explanation of decisions is an essential part of the decision itself, and lack of that makes the decision unacceptable. For example, a "robo-judge" deciding on a dispute between a customer and a health insurer is unacceptable without the explanation of the decision. This is referred to as the "Interpretability" problem. The book "Weapons of math destruction[7]" gives many interesting examples of this.

- Data privacy, transparency and control. All data and AI system exploit data, and many times this is personal data. Using all this personal data has as side effect that privacy may be compromised, even if it is unintentionally. The recent scandal of Cambridge Analytica / Facebook shows that this is a bigger issue than we might have thought[8,9]

To avoid those effects, people sometimes refer to the need for FATE AI (Fair, Accountable, Transparent and Explainable Artificial Intelligence).

2. Liability.

When systems become autonomous and self-learning, accountability of behaviour and actions of those systems becomes less obvious. In the pre-AI world, incorrect usage of a device is the accountability of the user, while device failure is accountability of the manufacturer. When systems become autonomous and learn over time, some behavior might not be foreseen by the manufacturer. It becomes therefore less clear who would be liable in case something goes wrong. A clear example of this are driverless cars. Discussions are ongoing whether a new legal person needs to be introduced for self-learning, autonomous systems, such as a legal status for robots[10], but it is generating some controversy[11].

3. Unknown consequences of AI.

The positive aspects of AI may have some consequences of which we don't know yet how they will work out.

- AI can take over many boring, repetitive or dangerous tasks. But if this happens at a massive scale, maybe many jobs might disappear and unemployment will skyrocket[12]?

- If less and less people work, then governments will receive less income tax, while costs of social benefits will increase due to increased unemployment. How can this be made sustainable? Should there be a "robot tax"[13],[14],[15]? How to be able to pay pensions when increasingly less people work?

- Is there a need for a universal basic income (UBI) for everybody[16]? If AI takes most of the current jobs, what do all unemployed people then live from?

4. How should people relate to robots? If Robots become more autonomous and learn during their "lifetime", then what should be the (allowed) relationship between robots and people? Could one's boss be a robot, or an AI system[17]? In Asia, robots are already taking care of elderly people, accompanying them in their loneliness[18],[19],[20]. And, could people get married to a robot[21]?

5. Concentration of power and wealth in a few very large companies[22],[23]. Currently AI is dominated by a few large digital companies, including GAFAM[24] and some Chinese mega companies (Baidu, Alibaba, Tencent). This is mostly due to those companies having access to massive amounts of propriety data, which might lead to an oligopoly[25]. Apart from the lack of competition, there is a danger that those companies keep AI as proprietary knowledge, not sharing anything with the larger society other than for the highest price possible[26]. Another concern of this concentration is that those companies can offer high-quality AI as a service, based on their data and propriety algorithms (black box). When those AI services are used for public services, the fact that it is a black box (no information on bias, undesired attributes, performance, etc), raises serious concerns, like when the LA Police Department announced that it uses Amazon's face recognition solution (Rekognition) for policing[27],[28]. The Open Data Institute in London has started an interesting debate on whether AI algorithms and Data should be closed, shared or open[29].

6. Intentional bad uses. All the points mentioned above are issues because AI and Data are applied with the intention to improve or optimize our lives. However, like any technology, AI and Data can also be used with bad intentions[30]. Think of AI-based cyberattacks[31], terrorism, influencing important events with fake news[32], etc.

There is an additional issue that also requires attention, which is the application of AI for warfare and weapons, especially for lethal autonomous weapons systems (LAWS). We do not discuss this issue here as this usually implies an explicit (political) decision, and is not something that will come as a surprise. Moreover, some will consider this good use, while other might call it bad use of AI. Some organizations are working on an international treaty to ban "killer robots"[33]. The issue recently attracted attention due to Google employees sending a letter to their CEO questioning Google's participation in defense projects[34]. This action may have contributed to Google publishing its AI principles stating the objectives for assessing AI applications and what AI applications they will not pursue[35].

3. Approach to move forward

While there is ample debate ongoing about many of those issues (e.g. AI and the future of work, today it is unclear what solutions there will be for some of the challenges. What we can define, however, are relevant actions to work on as an approach for dealing with those challenges. While executing the approach, adaptions can and will be made in a learning by doing process.

- Governments and institutions need to think about strategies and approaches to identify the issues along with their solution directions. The GDPR[36] is a small, but important step into that direction. Several national governments are already working on this through multidisciplinary committees of experts (see notes 2,3).

Maybe governments should ensure the availability of rich and sufficiently varied open datasets to minimize algorithmic bias.

- Private enterprises need to start thinking about self-regulation and about where they stand. They should be clear on how responsible they want to act and become.

- Probably a one-size-fits-all approach will not work, as some AI and Big Data applications have less potential negative side effect than others. E.g., decisions in marketing have probably less negative side effects than decisions about insurance premiums. Decisions made in so-called "safety critical" systems may need to be validated and verified by formal mathematical procedures to ensure their correct functioning under all possible circumstances[37].

- GDPR is a step forward regarding protection of personal data. A clear distinction needs to be made between Data & AI applications using personal data versus those using aggregated, anonymized data. Applications using personal data will need an explicit and transparent user consent (as provisioned in the GDPR). This is not needed when aggregated, anonymized data is used, but, as a matter of transparency, one might argue for the "right to be informed" when users' (aggregated, anonymized) data is used for applications.

- AI and Data should not only be used for commercial opportunities, but also for social opportunities, such as Data for Good and AI for Good[38],[39],[40], initiatives whose aims are to support achieving the United Nations Sustainable Development Goals (SDGs)[41].

- The Open Data[42] approach should be extended to Open Algorithms[43], to enable the benefits of private data, while not increasing the privacy risk and the commercial risks of private enterprises[44]. In this approach, algorithms are sent to the data, so data remains in its original premises, rather than the usual other way around, where data is stored centrally and then algorithms run on the data.

- Code of conducts should be in place for all professional families that contribute to Data and AI applications. This should reduce the likelihood of bad uses and unintended negative side effects.

- International, multidisciplinary committees should be put in place to oversee and monitor the uses of Data and AI across the world and raise alerts when needed. Maybe something analogue to the Civil Aviation Authority (for airplane crashes) could work.

4. A closing word

It is indeed important that we are prepared to ensure that our societies and economies continue to function well with the expected massive uptake of Data & AI applications. But we should neither forget that even without this, we don't live, and have never lived, in an ideal world. Think of the large number of humans that have taken, are taking and will take extremely wrong decisions, with hugely negative consequences for humanity. And many decision makers have taken -with good intent- important measures that have had serious negative side effects. So, while there are risks associated with the massive uptake of AI & Data, there is probably more to win than to lose with those technologies. Moreover, humanity has ample experience in how to manage or recover from negative consequences of (wrong) decisions.

If there is a final deadline before we would need to have solved the six challenges, it will probably be the Singularity[45], the point in time when Artificial Intelligence will lead to machines that are smarter than human beings. However, even if

this point may never arrive, it is good that societies are discussing the issues now and agree on a common vocabulary and framework, rather than waiting until it is (too) late.

---

[1] (2016, 12). Surviving the AI Hype-Fundamental concepts to understand Artificial Intelligence. *SlideShare*. 05, 2018,https://www.slideshare.net/LUCA-D3/surviving-the-ai-hupe-fundamental-concepts-to-understand-artificial-intelligence-70400058

[2] (2018, 03). AI in the UK: ready, willing and able?. *Parliament UK*. 05, 2018, https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

[3] (2017, 01). Rapport de synthèse France Inteligence Artificielle. *economie.gouv.fr*. 05, 2018, https://www.economie.gouv.fr/files/files/PDF/2017/Rapport_synthese_France_IA_.pdf

[4] (2017, 04). Just like humans, artificial intelligence can be sexist and racist. *Wired*. 05, 2018, http://www.wired.co.uk/article/machine-learning-bias-prejudice

[5] (2018, 03). How to Prevent Discriminatory Outcomes in Machine Learning. *World Economic Forum*. 05, 2018,https://www.weforum.org/whitepapers/how-to-prevent-discriminatory-outcomes-in-machine-learning

[6] (2018, 01). Confusion matrix. *Wikipedia The Free Encyclopedia*. 05, 2018, https://en.wikipedia.org/wiki/Confusion_matrix

[7] Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy, Crown Publishing Group New York, NY, USA ©2016, ISBN:0553418815 9780553418811

[8] (2018, 03). Cambridge Analytica The Cambridge Analytica Files Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*. 05, 2018,https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

[9] (2017, 03). How Trump Consultants Exploited the Facebook Data of Millions. *The New York Times*. 01, 2018,https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html

[10] (2017, 01). MOTION FOR A EUROPEAN PARLIAMENT RESOLUTION. *European Parliament*. 05, 2018, http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+REPORT+A8-2017-0005+0+DOC+PDF+V0//EN

[11] (2018, 01). OPEN LETTER TO THE EUROPEAN COMMISSION ARTIFICIAL INTELLIGENCE AND ROBOTICS. *Robotics Openletter*. 05, 2018, http://www.robotics-openletter.eu/

[12] (2017, 05). Technology, jobs, and the future of work. *mckinsey&company*. 05, 2018, https://www.mckinsey.com/global-themes/employment-and-growth/technology-jobs-and-the-future-of-work

[13] (2017, 03). Robots won't just take our jobs – they'll make the rich even richer. *The guardian*. 05, 2018, https://www.theguardian.com/technology/2017/mar/02/robot-tax-job-elimination-livable-wage

[14] (2017, 02). The robot that takes your job should pay taxes, says Bill Gates. *Quartz*. 05, 2018, https://qz.com/911968/bill-gates-the-robot-that-takes-your-job-should-pay-taxes/

[15] (2017, 02). European parliament calls for robot law, rejects robot tax. *Reuters*. 05, 2018, https://www.reuters.com/article/us-europe-robots-lawmaking-idUSKBN15V2KM

[16] 017, 12). We could fund a universal basic income with the data we give away to Facebook and Google. *The next web*. 05, 2018, https://thenextweb.com/artificial-intelligence/2017/12/05/we-could-fund-a-universal-basic-income-with-the-data-we-give-away-to-facebook-and-google/

[17] (2016, 09). Why a robot could be the best boss you've ever had. *The guardian*. 05, 2018,https://www.theguardian.com/media-network/2016/sep/09/robot-boss-best-manager-artificial-intelligence

[18] (2017, 08). Robot caregivers are saving the elderly from lives of loneliness. *Engadget*. 05, 2018, https://www.engadget.com/2017/08/29/robot-caregivers-are-saving-the-elderly-from-lives-of-loneliness/

[19] (2018, 02). Japan lays groundwork for boom in robot carers. *The guardian*. 05, 2018,https://www.theguardian.com/world/2018/feb/06/japan-robots-will-care-for-80-of-elderly-by-2020

[20] (2018, 03). HOW ROBOTS WILL FIND YOU IN RETIREMENT HOMES AND TAKE YOU TO BINGO: A VISION FOR PERSON SEARCH AND TASK PLANNING. *IEEE Intelligent Systems*. 05, 2018,https://publications.computer.org/intelligent-systems/2018/03/27/robots-retirement-homes/

[21] (2017, 04). Chinese man 'marries' robot he built himself. *The guardian*. 05, 2018, de https://www.theguardian.com/world/2017/apr/04/chinese-man-marries-robot-built-himself

[22] (2018, 01). Report: AI will increase the wealth inequality between the rich and poor. *AI News*. 05, 2018,https://www.artificialintelligence-news.com/2018/01/11/ai-wealth-inequality/

[23] (2016, 12). AI could boost productivity but increase wealth inequality, the White House says. *CNBC*. 05, 2018,https://www.cnbc.com/2016/12/21/ai-could-boost-productivity-but-increase-wealth-inequality-the-white-house-says.html

[24] Google, Amazon, Facebook, Apple, Microsoft

[25] 2016, 11). AI could boost productivity but increase wealth inequality, the White House says. *CNBC*. 01, 2018,https://siliconangle.com/blog/2018/03/01/democratized-ai-could-put-the-tech-oligopoly-in-its-place-ibmml/

[26] (2018, 02). GOOGLE MUST BE STOPPED BEFORE IT BECOMES AN AI MONOPOLY. *wired*. 05, 2018, https://www.wired.com/story/google-artificial-intelligence-monopoly/

[27] (2018, 05). Amazon is selling police departments a real-time facial recognition system. *The Verge*. 01, 2018, https://www.theverge.com/2018/5/22/17379968/amazon-rekognition-facial-recognition-surveillance-aclu

[28] (2018, 05). Amazon Pushes Facial Recognition to Police. Critics See Surveillance Risk.. *The New York Times*. 05, 2018,https://www.nytimes.com/2018/05/22/technology/amazon-facial-recognition.html

[29] (2018, 04). The role of data in AI business models (report). *Open Data Institute*. 05, 2018, https://theodi.org/article/the-role-of-data-in-ai-business-models/

[30] (2018, 02). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. *Open Data Institute*. 05, 2018,https://img1.wsimg.com/blobby/go/3d82daa4-97fe-4096-9c6b-376b92c619de/downloads/1c6q2kc4v_50335.pdf

[31] (2018, 02). Artificial intelligence poses risks of misuse by hackers, researchers say. *Reuters*. 05, 2018,https://www.reuters.com/article/us-cyber-tech/artificial-intelligence-poses-risks-of-misuse-by-hackers-researchers-say-idUSKCN1G503V

[32] (2018, 04). Artificial intelligence is making fake news worse. *Business Insiders*. 05, 2018,http://uk.businessinsider.com/artificial-intelligence-is-making-fake-news-worse-2018-4

[33] (2018, 04). France, Germany under fire for failing to back 'killer robots' ban. *Politico*. 05, 2018,https://www.politico.eu/article/artificial-intelligence-killer-robots-france-germany-under-fire-for-failing-to-back-robots-ban/

[34] (2018, 04). 'The Business of War': Google Employees Protest Work for the Pentagon. *The New York Times*. 05, 2018, https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html

[35] (2018, 06). AI at Google: our principles. Google blog. 06, 2018 https://blog.google/topics/ai/ai-principles/

[36] (2018, 01). The EU General Data Protection Regulation (GDPR) is the most important change in data privacy regulation in 20 years - we're here to make sure you're prepared.. *EUGDPR.org*. 05, 2018,https://www.eugdpr.org/

[37] (2007, 04). Formal Modeling and Verification of Safety-Critical Software. *IEEE Xplore Digital Library*. 05, 2018,https://ieeexplore.ieee.org/document/4814957/

[38] 2018, 04). AI for Good Global Summit 2018. *IEEE Xplore Digital Library*. 05, 2018,https://www.itu.int/en/ITU-T/AI/2018/Pages/default.aspx

[39] (2018, 01). Artificial Intelligence for Sustainable Global Development. *AI for Good Foundation*. 05, 2018,https://ai4good.org/

[40] (2016, 10). Big Data for Social Good. LUCA Data-Driven Decisions. 05, 2018, https://luca-d3.com/big-data-social-good/index.html

[41] (2016, 01). Sustainable Development Goals. UNITED NATIONS DEVELOPMENT PROGRAMME. 05, 2018, http://www.undp.org/content/undp/en/home/sustainable-development-goals.html

[42] (2012, 01). The Open Data Institute. *The ODI*. 05, 2018,https://theodi.org/

[43] (2017, 01). Open Algorithms for better decisions. *OPAL*. 05, 2018,http://www.opalproject.org/

[44] (2017, 01). Developing privacy-preserving identity systems and safe distributed computation, enabling an Internet of Trusted Data.. *The Trust::Data Consortium*. 05, 2018,https://www.trust.mit.edu/

[45] (2015, 05). Technological singularity. *Wikipedia, The Free Encyclopedia*. 05, 2018, https://en.wikipedia.org/wiki/Technological_singularity

About LUCA

LUCA is Telefónica's specialist data unit, which sits within the Chief Data Office, led by Chema Alonso. Its mission is to bring Telefónica's know-how in transforming into a data-driven organisation to other private and public sector organisations in a wide range of sectors including Retail, Tourism, Outdoor Media, Financial Services and Transport. Its diverse product portfolio, which brings together expertise in Artificial Intelligence, Data Engineering, Data Science and Infrastructure, enables companies to accelerate their Big Data journey with a wide range of solutions and expertise to propel their digital transformation.

More information

luca-d3.com

@LUCA_D3