

## Artificial Intelligence its Potential and Limits

### Introduction

Underestimating the societal impact of the (emerging) Artificial Intelligence technologies would nowadays be, to say the least, frivolous. In this vein, plenty of initiatives have already put in place around the globe, aimed at understanding and subsequently designing and deploying appropriate action-programs needed to address the societal challenges posed by Artificial Intelligence within the next few years [1].

The current issue, presents a number of selected contributions which analyze the impact of Artificial Intelligence on key societal sectors, with an eye on the initiatives already put in place in the Basque Country and Navarre.

Artificial Intelligence (AI for short) comprises a rather fuzzy demarcation for an emerging field which is claimed to carry along a disruptive technology applicable for almost every aspect of ordinary (and even extraordinary) human life, and perhaps even of death too. AI enables machines to respond in human-like ways by (machine) learning from previously accomplished tasks in order to adjust themselves to new external stimuli. This involves processing of large amounts of data (big data) and recognizing patterns in the data. This combined technology, namely, AI powered by machine learning, has led to a disruptive innovation which is foreseen to impact many global trends in the near future. These expand from scientific discovery itself to efficient technology implementation with large implications in societal issues including health-care, manufacturing, banking and financing, retail managing, etc.

The origin of AI constitutes nowadays a highly debated issue, which will not be discussed here. Let us simply note that 50 years ago, Herbert Simon [2] wrote a masterpiece book which brought AI to the fore of the scientific arena. He envisaged AI as the scientific study, and usage- of “intelligent machines”, regarded them not as simple gadgets, but as a class of stakeholders with distinctive behaviors, able to create ecological environments, either by themselves or in association with humans.

After its fifty years of existence, AI has notably start to change our lives, for now we can process images automatically, segment them at will and provide semantic descriptions of their content, certainly an intelligent behavior. In this vein, AI is challenging some human intelligent behaviors like those of the best medical doctors at diagnosing diseases like cancer, or detecting regularities and “connectivity patterns” in the brain, as well as anomalies in a myriad of radiological images. AI can also drastically increase the reliability of clinical trials offering tools for finding out eligible patients. For instance, an estimated 20% of people suffering from cancer are eligible for most clinical trials, but on average less than 5% take part [3], simply because they rest of them remain unidentified.

AI technologies powered with machine learning do provide nowadays digital therapies for common physical and mental maladies of the elder [4] that have notably improved their life's quality. Additionally, autonomous surgical robots, trained by AI have already made their debut in surgery rooms with successes paralleling those of the best surgeons in many precision surgical tasks. As their usage increases, it is reasonable to expect that their performance will

also increase, as these AI governed robots (machine) learned from their own data as it is accumulated in each execution of the code.

These are a few of the spectacular achievements of AI in medicine. See Ref [5], though, for a more complete review.

However, the irruption of AI in modern medicine has concomitantly raised a number of concerns, related to the bias encoded in the training data, and to the privacy of data itself. The former concern follows from the fact that the training data is the recorded data, and this corresponds to segments of population that have traditionally had access to medical attention, putting aside features pertaining to that population segment which has suffered from less access to medical care. The second concern stems from the fact that keeping privacy of medical data is difficult, even in developed countries. For instance, the University of Chicago has recently faced a lawsuit for sharing medical data with Google. Naturally, the University of Chicago stripped all kind of identifications, like names, social security numbers, etc. from the data, but not the dates of patients' visits. The point in the lawsuit was that Google could restore that data by combining them with information already held by Google, like smartphone locations, to assign names to medical records.

Finally, it is also worth mentioning that many doctors have come to hate their computer terminals because they must now spend more time feeding data into the system than interacting with their patients.

Machine vision and scenario analysis is paving the road toward vehicle's autonomous driving devices. If such an initiative materializes, within a few years' time from now, as predicted [6], it will shape in nowadays unimaginable manners the mobility of people and consumer goods around the globe. Analogously, since modern big data technology allows to store the ~2.5 quintillion ( $2.5 \times 10^{18}$ ) bytes of data generated by people each day, machines powered with AI are being used to harness these data in order to provide specialized services to retailers aimed at putting their products in the market in the most efficient (productive) way [7]. Additionally, consumer goods and services' prices are increasingly set by AI software, operating by on-line platforms, voters receive nowadays more often than ever, targeted, biased (political) fabricated information through social media, thus threatening the very foundations of democracy. In the same vein, it is worth noticing that the combined power of high-quality facial image recognition and scenario analysis provide the tools of surveillance [8] which could be used to control individuals as anticipated by George Orwell in his classic futuristic novel [9], written more than 70 years ago.

However, AI powered with machine learning could bring enormous benefits if used properly, for like it or not, the digital realm is where citizens live and work, shop and play, meet and fight. Consequently, for responsive governments, ensuring democratic rules, establishing strategic and tactical future scenarios and providing funding for all of these, requires understanding of big data and AI algorithms. Notice that the data that the government collects from and about their citizens could, in principle, be used to tailor education to the needs of each child, fit health care to the lifestyle, perhaps also to the genetics of each patient, predict traffic deaths, street law offense, natural disasters, planning personalized care for the elder, etc. Clearly, technological innovations based of AI are becoming essential for the responsive government to keep its position of democratic authority in a data-intensive world. But, a word of caution is in order here, for trust in the public sector is severely eroded when projects fail, a fact that brings to the fore, the importance of fair and expert advice to

the government. In this respect it has been recently put forward that independent academic researchers are better placed than (large) companies to help governments to maximize the social return of the use of AI technologies [10, 11].

One key issue in this respect is the cybersecurity of the data captured, stored and utilized by the public institutions, in particular, and by the remaining organizations, in general. Artificial Intelligence powered by machine learning is increasingly being used for verifying codes and identifying bugs and vulnerability. One obvious procedure for this is to spot abnormal activity, shutdown in-out access to such part of the network and issue and alert message. In this way, AI makes deterrence possible because cyberattacks can be identified and neutralized before the agent causing them is located. The value of AI powered with machine learning technologies in cybersecurity was estimated to amount  $10^9$  Euro in 2017, and predicted to reached  $15 \times 10^9$  Euro by 2023 [12]. Progress on these matters needs to address urgently the distinction between legal and illegal actions, as well as the considerations of “sparring” exercises between institutions to test AI based cybersecurity protocols, before they are put in place. Recall that AI learns by gained experience in a great deal, hence, these exercises will help to improve the security protocols. Last but not least, the unexpected must be expected to occur, i.e.: unintended human mistakes, technical failures, etc. and be prepared to address them in the most efficient way.

We can also process natural languages by software based on artificial intelligence techniques allied with the scientific study of languages. Though it is widely acknowledged that we are still far away from full machine understanding of natural language, for automated translations still need to be reviewed and edited by skilled human translators while no computer system has yet come close to passing the ‘Turing Test’ of convincingly simulating human conversation, many websites now offer automatic translation, mobile phones can appear to understand spoken questions and commands, search engines use basic linguistic techniques for automatically completing or ‘correcting’ queries and for finding relevant results that are closely matched to the searched terms, in such a way that natural language processing has nowadays leave universities and research laboratories to inform a variety of industrial and commercial applications [13].

The impact of AI technologies on modern large techno-social systems is also foreseen to be massive. Such large-scale infrastructures, like global transportation power and gas distributions grids, the internet itself, are nowadays nested in a global web of communication and computing interconnected machines, whose systemic dynamics and evolution is beyond human comprehension, though they are ultimately determined by human (controller) actions. Forecasting the phenomena in large techno-social systems is hampered by the limited knowledge and the overwhelming complexity of the myriad of interactions among their countless components, and the “unpredictable” human behavior and societal factors at stake, like for instance, the disruption of social order during emergencies such as pandemics. To effectively control and predict the performance of such systems, it is necessary to have a clear picture of the patterns emerging in real-world data, which could subsequently be used to anticipate failures and eventually catastrophic events, evaluate risks, and assess the quality of the services provided [14].

The interactions of humans with AI powered machines can be divided into those mediated by either signals or speech communication with machine, and those where a certain kind of symbiosis with the machine is built. The latter has in itself three further categories, the one

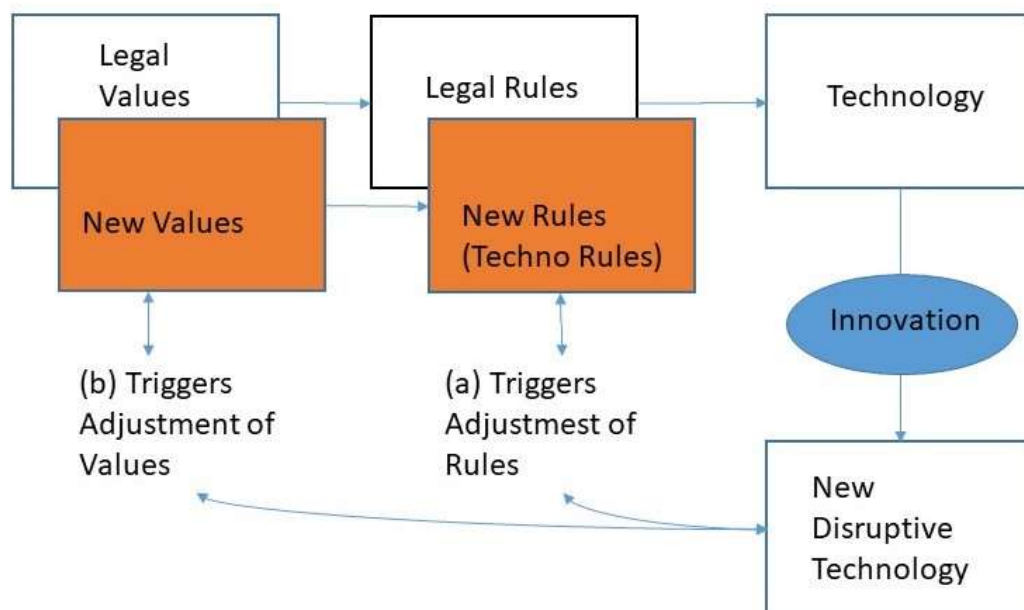
in which the machine records the activity of the brain and decodes its meaning, the one in which the machine stimulates the brain to manipulate its activity and affects its functions, and the one in which both above mentioned activities occur simultaneously, the latter is known as “brain-machine bidirectional coupling”. In this realm, signals recorded from the brain’s activity are increasingly being (pre)processed by AI software before specific commands are delivered to the prostheses. For instance, such an approach has been applied to process the neural activity occurring when people with epilepsy mouths words, which subsequently are delivered to a speaker to generate synthetic speech [15]. In spite of how spectacular these achievements might be, they introduce serious ethical concerns, as AI software learns from both the training data and the supplied data to generate algorithms that cannot be fully traced down, and hence are impossible to comprehend [16]. This puts a big unknown between a person’s thoughts and the actions of the prosthesis which operates on his behalf [17], thus giving rise to ethical issues about the intentionality of these “human” acts, and the capability of humans to act freely and in accordance with their own will.

With respect to the first category of human-AI powered machine interactions, mentioned in the preceding paragraph, the key issue is how trust is attributed by humans to machines and vice versa. Surprisingly, there is mounting evidence that humans tend to trust more advice coming from AI algorithms than from their fellows, and indeed, it has been recently claimed that such behavior is more pronounced, the tougher is the decision to be made [18]. For attributing trust in the other direction, namely from machines to human’s bold steps have begun to be taken towards building robust AI algorithms for artificial cognitive architectures. Developmental robotics paradigms which can estimate the trustworthiness of its human interactors which have recently been explored with some success [19], seem to be promising.

AI powered with machine learning is also foreseen to have large impact in human decision-making tasks. In particular, AI nowadays already supports many decisions made by judges in court. Consider the millions of times per year where judges, after arrest of the defendant, decide whether she/he should await trial at home or in prison. These decisions are based on a “reasonable” guess make by the judge of what would do the defendant if she/he were released. Given the large amount of data relative to such cases available throughout many judiciary repositories, makes AI powered with machine learning a promising tool for this task, because in this case “predictions” could be made based on a huge amount of data properly processed by AI algorithms which are able to unveil hidden patterns not accessible at all from the few cases that each judge can keep in her/his head. It does not come as a surprise, though, that for many, the prospect of robot-judges seems plausible (even imminent, for some). The implicit move towards a “codified justice” rather than discretionary moral judgment ruling nowadays is also welcomed by the advocates of implementing AI in court procedures. The increasing usage of AI in court will predictably reinforce values associated with codified justice, and that is likely to push for a self-reinforcement for further usage of AI which will raise a number of concerns related with the *incomprehensibility*, namely the impossibility of humans to traced down AI algorithms, the *datafication*, for emphasis on recorded data can deprive the legal system form legitimate criticism, giving bias a change to flourish (vide infra). For instance, AI algorithms in use in New York City have been trained with allegedly biased data for these data cannot, by construction, account for whether imprisoned defendants would have committed crimes had they been released, simply because they had never been. Consequently, the training data reflects the preferences and prejudices of the judges that have made such decisions in the past, preferences that will remain in place with the so-trained AI decision making system too.

Two sensible “nontechnical issues” refer to the *disillusionment* which widespread use of AI can project on earlier practices plagued with multitude failures made of human judges, which will pose into question the legitimacy of the legal system, and to the *alienation* of human participation for it will promote the cease of humans’ participation, or at least the loss of interest, in its operations [20].

However, the irruption of a technology like AI could also catalyze changes in the law system that need to be considered beforehand. Two main paradigms have been put forward, based on a (over)simplified model of the justice adjudication procedure shown in the Figure below.



The path marked as (a) depicts the effect that the introduction of the “new disruptive technology” may exert on the system affecting only the adjustment/substitution of “Legal Rules”, but keeps the values on which such rules are based upon unaffected. However, more profound changes could be triggered affecting to the values themselves, as outlined by the path marked with (b).

Another implication of the irruption of AI technologies in court procedures refers to the foreseen erosion of the “prestige of experts”, a value deeply rooted in the justice administration system, as it will be replaced, only partly perhaps, by the decisions made by the AI powered machine. Needless to say, that the same applies to the witnesses and the members of the jury alike. They will see their roles largely limited. These “radical” changes need to be taken in serious consideration and levered accordingly before it is too late.

AI powered with machine learning has already made a massive impact in modern synthetic chemistry [21], molecular biology [22], and materials discovery research [23]. Up to date the Schrödinger (time independent) equation,  $\hat{H}(\{Za, Ra\})\Psi = E(\{Za, Ra\})\Psi$ , has been used to set up a well-tested, accurate computational scheme for structure-property relationships of molecules and materials. The former, the structure, is determined by the identity, nuclear charges  $\{Za\}$ , of its constituent nuclei, and their positions,  $\{Ra\}$ . The latter, the properties,  $\{O\}$ , obtained from the wave function,  $\Psi$  as  $O = \langle \Psi | \hat{O} | \Psi \rangle$ , being  $\hat{O}$  the quantum mechanical operator of property  $O$ . Most modern structure-property simulation toolkits allow the sought properties to be anticipated with high accuracy even before the compound has been made.

But techniques based on AI powered with machine learning have the potential to dramatically change and enhance the role of computation in materials synthesis and discovery research by introducing an approach based on pattern search and recognition from big data repositories, which can be carried out in an efficient manner by AI searching of massive combinatorial spaces or complex nonlinear processes. The former concerns mainly with material discovery research, and the latter constitutes the salient feature of the designing of synthetic procedures. Notice, nonetheless, that at the heart of these AI powered with machine learning techniques lie a number of statistical algorithms whose performance improves as better training is exercised, and more data is supplied, following the law of large “uncorrelated” data sets whose uncertainty (variance of the variables of data set) decreases as  $N^{-1/2}$ , for large enough  $N$ . However, this assumption should not be taken for granted. There is no reason to claim that the data sets used in materials discovery research are “uncorrelated”. Consequently, a word of caution is in order at this point [24].

All in all, AI powered with machine learning technologies will change drastically our world, both, physically and virtually. Managing the changes generated by the application of these technologies to various realms of human activity, requires an understanding of the essentials of the technology and a realistic prospective view of the changes it could produce.

This issue aims at addressing such requirements by providing an overall view of the potential and the limits of artificial intelligence.

#### References:

- [1] It is worth mentioning that the European Union has recently set up “The Committee on Artificial Intelligence in the Digital Age (AIDA)”, and gave it the mandate to present a EU “roadmap” that encompasses the steps the European Union needs to take in order to balance obvious benefits and potential risks driven by Artificial Intelligence technologies on health, infrastructure, sustainability, transport, agriculture, energy, defense, industry, democracy, e-government, employment, skills and education. See: Axel Voss, *Report on Artificial Intelligence in a Digital Age* ([2020/2266\(INI\)](#)), Document: A9-0088/2022, Special Committee on Artificial Intelligence in a Digital Age, (2022).
- [2] H. A. Simon, *The Sciences of the Artificial*, MIT Press, Cambridge, USA, (1969)
- [3] T. C. Haddad et al. *J. Clin. Oncol*, 36, 6550 (2018)
- [4] S. Makin, *A Smarter Way to Treat*, *Nature*, 573, S106 (2019)
- [5] Nature Outlook, *Digital Health. Transforming Medicine with Artificial Intelligence*, *Nature*, S98-S116 (2019)
- [6] A. Nunes, B. Reimer, J. F. Coughlin, *People Must Retain Control of Autonomous Vehicles*, *Nature*, 556, 169 (2018).
- [7] K. Letheren, R. Russell-Bennett, L. Whittaker, *Black, White or Grey Magic? Our Future with Artificial Intelligence*, *J. Marketing Manag.* 36, 216-232 (2020).
- [8] S. Zuboff, *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power.* (2019).
- [9] G. Orwell, *Nineteen-Eighty-Four*, Secker & Warburg, London (1950).

- [10] Y. Benkler, *Don't Let Industry Write the Rules of AI*, *Nature*, 569, 161 (2019)
- [11] Nature Editorial, *Experts Must Demand to be Heard*, *Nature*, 572, 153 (2019)
- [12] P&S marker Research, *Artificial Intelligence in Cybersecurity Market*, (2017)
- [13] D. Jurafsky, J. H. Martin, *Speech and Language Processing: An introduction to speech recognition, computational linguistics and natural language processing*, Prentice Hall, Upper Saddle River, NJ, USA (2008)
- [14] A. Vespignani, *Predicting the Behavior of Techno-Social Systems*, *Science*, 325, 435-428 (2009)
- [15] G. K. Anumanchipalli et al. *Nature*, 568, 493-498 (2019)
- [16] AI is impossible for humans to trace down to the point of being in a position to justify every single bit of the computational sequence that produces the output. See: P. Humphreys, *The Philosophical Novelty of Computer Simulation Methods*, *Synthese*, 169, 615-626 (2009)
- [17] L. Drew, *Agency and the Algorithm*, *Nature*, 571, S19 (2019)
- [18] E. Bogert, A. Schechter, R. T. Watson, *Humans Rely More on Algorithms than Social Influence as a Task Becomes More Difficult*, *Sci. Rep.*, 11, 8028 (2021)
- [19] S. Vinanzi et al. *Would a Robot Trust You? Developmental Robotics Model of Trust and Theory of Mind*. *Phil. Trans. R. Soc. B.* 374, 20180032 (2018)
- [20] R. M. Re., A. Solow-Nierderman, *Developing Artificial Intelligent Justice*, *Stan. Tech. L. Rev.* 22, 242 (2019)
- [21] P. M. Pflüger, F. Glorius, *Molecular Machine Learning: The Future of Synthetic Chemistry*, *Angew. Chem. Int. Ed.* 59, 18860 (2020)
- [22] E. Callaway, *AI makes Gigantic Leap in Solving Protein Structures*, *Nature* 588, 203 (2020)
- [23] J. Westermayr, M. Gastegger, K. T. Schütt, R. J. Maurer, *Perspective on Integrating Machine Learning into Computational Chemistry and Materials Science*, *J. Chem. Phys.* 154, 230903 (2021)
- [24] S. Succi, P. V. Coveney, *Big Data: The End of the Scientific Method?* *Phil. Trans. R. Soc. A* 377, 20180145 (2019)

Javier Echeverría<sup>1</sup> and Jesus M. Ugalde<sup>2</sup>

---

<sup>1</sup> Jakiunde, Zientzia, Arte eta Letren Akademia. Prim, 7 – E-20006 – Donostia-San Sebastián. Euskal Herriko Unibertsitatea (UPV/EHU)

<sup>2</sup> Jakiunde, Zientzia, Arte eta Letren Akademia. Prim, 7 – E-20006 – Donostia-San Sebastián. Euskal Herriko Unibertsitatea (UPV/EHU). Donostia International Physics Center, DIPC. Manuel Lardizabal, 4. E-20004 - Donostia-San Sebastián