

Big Data & Machine Learning Data Analytics & New Trends

Alfredo M. Ronchi
alfredo.ronchi@polimi.it

Artificial Intelligence (AI) together with its sister-tech Machine Learning (ML)¹ and Big Data (BD) open great opportunities in many fields of public interest: education, training, health, safety and security, public services, as well as for democratic processes, political decision-making, and civic participation.

We would like to identify the role and impact of AI and Big Data, both private and public uses, in addition to clear benefits, contain some risks at the expense of citizens' rights and chances for enhancing direct democratic participation. We will specifically consider AI and BD impact on individuals, on society, on institutions, and on the political process in the age of datafication and digital transition.

We will investigate and identify the use of AI and Big Data in the field of forecasting and citizen's sentiments, and the use of AI and BD in the field of political decision making, providing a transparent rationale about the decision-making process avoiding as much as "black boxes" thanks to an innovative methodology.

Advances in the development and adoption of Artificial Intelligence across a wide range of applications have confronted policymakers in many developed countries with the increasingly pressing need to update and amend their respective legislative and regulatory frameworks in response to the new challenges associated with this technology.

The EU has emerged as a leading player in the development of such frameworks. Already in April 2019, the High-Level Expert Group on AI presented Ethics Guidelines for Trustworthy Artificial Intelligence. Based on these Guidelines, the EU Commission, in its 2020 white paper on AI, and subsequently in the Artificial Intelligence Act², identified 7 major gaps that must be resolved for AI to become a safe, widely adopted technology that brings benefits without creating equal or larger risks or divides to operators, citizens³ and society:

- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination, and fairness
- Societal and environmental wellbeing
- Accountability and acceptance

¹ In the present document, where not strictly necessary, we will group under the term AI both technologies AI and ML.

² AI white paper https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

Artificial Intelligence Act <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>

Fostering a European approach to Artificial Intelligence <https://www.astrid-online.it/static/upload/comm/0000/communication-fosteringauropeanapproachpdf.pdf>

³ Alfredo M. Ronchi [2019] e-Citizens: Toward a New Model of (Inter)active Citizenry, ISBN 978-3-030-00745-4, Springer

Each of these gaps is a direct symptom of the black box problem in AI / ML and sometimes even in BD analysis, these gaps provide a good general description of the problem facing every stakeholder of AI systems across all domains.

AI and BD pose new challenges both to core individual values such as privacy, freedom, and equality, as well as to European collective values, such as fairness, security, inclusiveness, accountability, wellbeing, resilience and access to education and democratic control.

They touch on fundamental rights of EU citizens that are enshrined in the EU Charter of Fundamental Rights (the 'Charter') and protected by an extensive body of EU secondary, and member-states law.

The rights to privacy and data protection (Arts.7 and 8 of the Charter) are implemented and made enforceable in EU secondary law by the General Data Protection Regulation (GDPR). Art. 21 GDPR grants citizens a right to object against profiling based on personal data. while Art. 22 GDPR states that citizens have the right 'not to be subject to automated decision making, including profiling, which produces legal effects or otherwise significantly affects them and/or their personal interests.

Protections for EU citizens against discrimination on the grounds of sex, race, colour, language, political or religious views, and other personal or social attributes (Art. 21 of the Charter and Arts. 10 and 19 TFEU) come into focus every time algorithms are used to customise or personalise products and services, in particular when delivered through digital channels.

The 'black box' effect associated with current AI / ML models also raises issues regarding citizens' right to effective legal redress (Art. 47 of the Charter) with breaches of their rights becoming difficult to identify, and liability hard to attribute.

We must identify how to systematically protect fundamental rights and European values⁴ from possible threats stemming from the use of artificial intelligence (AI) and big data (BD) applications.

On the other hand, we must explore new pathways on how to foster the human centric use of AI and BD to enhance citizen easy access, engagement, and democracy⁵. In its Coordinated Plan on AI (2021 Review), the Commission has formulated the ambition to '*make the public sector a trailblazer for using AI*'. including e-Petitions and how to manage the changes of opinions in rapidly changing societal, natural, political and security environments. Design and promote new ways to reach awareness and educate European citizens about these technologies to enable informed civic participation in shaping them⁶. These objectives will be achieved thanks to an impact-oriented citizen's and human-AI interaction awareness campaign and capacity building in AI and BD. Within the scope of the proposed project will be included policy recommendations to improve data governance and promote the use of unbiased and resilient AI and BD. New and established regulations will be carefully considered to monitor platforms using these technologies (e.g. the previously mentioned AI Act, the Data Governance and Data Acts, the Digital Services⁷ and Markets Acts, GDPR, and other current and forthcoming legislative initiatives) leading to sustainable solutions protecting from the possible use, abuse and misuse negative impacts of these new

⁴ Information Self-Determination could be such a core value (referring to use case 1)

⁵ Alfredo M. Ronchi [2019], e-Democracy: Toward a New Model of (Inter)active Society, ISBN 978-3-030-01595-4, Springer

⁶ Awareness and education campaign must address EU citizens from kids to seniors.

⁷ Alfredo M. Ronchi [2019] e-Services: Toward a New Model of (Inter)active Community, ISBN 978-3-030-01841-2, Springer

technologies on fundamental rights and democracy. Existing frameworks and guidelines, such as the OECD Good Practice Principles for Data Ethics in the Public Sector, will be considered and EU-specific requirements further developed. Different use-cases will be developed within the proposal lifetime. Applications that are defined as “high-risk” in the EU’s AI Act⁸ will be given particular priority because organisations must become compliant with the ACT over the next few years. We will propose technology⁹ that can make them fully compliant easily, without disrupting their entire AI operations

The research work within the project will focus on the impacts of the following topics on democracy:

- Big Data key aspects: collection, anonymisation, pseudo anonymisation, GDPR compliance, filtering, etc.
- Big Data Analytics: big data lakes, advanced analytical tools, patterns extraction,
- Investigating current machine learning and AI models used in the political and civic environments and their current practices,
- Building workflows for better AI “explainability” of existing systems,
- Creating a framework to allow audience specific awareness campaigns and education modules (for example for government officials, for citizens, for external auditors of the democratic processes),
- Building use-cases and running pilots on local citizen participation actions, local elections, referenda, polls, surveys etc.,
- Foster participatory design of democratic vision and processes when building trustworthy intelligent infrastructures, e.g., for e-voting or participatory e-budgeting,
- Creating policy recommendations for local, regional, national, and European levels to integrate “explainability” of AI systems to safeguard democracy.

Detailed description of the research work and specific areas of investigations for each of the topics will be included in the paragraphs devoted to related tasks. The consortium includes a mix of technical, legal, and social science experts, coming both from Academia, Public Bodies, private sector, together with local administrations and civil society groups that are the ones that will be mostly impacted if actions are not taken now. Public authorities are also invited to participate as either full or associated partners to take part in the use-cases.

Impact on society

Digital technology in general had and still have a strong impact on society and the pandemic accelerated and amplified such impact especially on young generations. Leveraging on laziness and relaxation citizens spend less time outside home, they have shopping online, they buy food and drinks directly delivered on their table, “meet” friends on Zoom or WhatsApp, interact with the “outer environment” though the mediation of social media and video clips. These aspects are even more evident in young generations that add to the social media the gaming dimension. Of course, such trends are even amplified by other media such as television and news.

⁸ Annex III of the AI Regulation does not really specify organisations but lists 'high-risk' areas of application. The scope of these applications is quite wide.

⁹ e.g., provide to AI modules a digital identity and a wallet to host certificates - mixing existing AI technology and wallet technology into "Innovative uses of AI and big data".

Impact on privacy

People use to think that cyberspace is a “black hole” without memory where you pour data without any side effect. Young generations shared online sensitive information to access a videogame or chat with friends or more recently posted images and clips about their private life.

What about the push message asking to provide details about your activities yesterday evening, something that your digital “buddy” was unable to trace? Your bank will suggest, accordingly with some intelligent algorithms the average monthly expenses due to profiles matching with yours and send an alert if you are exceeding the limits. Computer vision will enable your smartphone to identify every single person in a group you photographed and video analysis plus 3D real-time modelling enable intelligent optimisation algorithms to improve human performances, wearable sensors and IoT complete the schema.

Impact on security

We all know that security and privacy are subject to risk as already stated thus it is important to identify and mitigate risks associated with privacy and security concerns. In order to reach this goal, as a first approach, we can perform the following steps: identify the persons at risk in the event of personal information exposure (not restricted to the data owner or collector), identify knowledge assets that can be extracted from the data collected (discrete data points, meta-analysis of data points, mash up of the collected data and external data sources); evaluate the importance of each knowledge asset to the potential goals/harms (little or no relevance, significant relevance, crucial).

Impact on decision making

The extensive use of AI, ML and Big Data, apart from several ethical issues, can led to some relevant drawbacks. As an example, let’s consider “nudging”.

The concept of nudge is already used in digital systems even if the nature of the mechanisms that characterise it is not always consistent, and some uses overflow into practices already prohibited by current legislation. In fact, the use of even “slight” and often morally irrelevant manipulations of the architecture of the decision is constrained both in the use of personal data to be able to construct a nudge mechanism (by the GDPR) and if the desired result falls within the category of fraudulent transactions (thanks to the UCPD¹⁰). The progress of AI has made it possible to develop much more powerful nudge mechanisms thanks to the effectiveness of statistical and inferential AI systems. The impact of AI powered technology on human autonomy is huge. AI-enhanced nudges reinforce the ability to achieve the designer goals using cognitive biases, emotional impulses, and other human behavioural mechanisms both intentionally and unintentionally.

Impact on opinion dynamics in social networks

Opinion formation is a complex and dynamic process mediated by interactions among individuals in social networks, both offline and online. Social media have drastically changed the way opinion dynamics evolve, in any case, they provide a reservoir of data for the study of opinion dynamics on social networks. Social media have become a battlefield on which opinions are, often violently, exchanged. In turn the behaviour of social media has become an important early indicator of societal change.

¹⁰ Unfair commercial practices directive https://ec.europa.eu/info/law/law-topic/consumer-protection-law/unfair-commercial-practices-law/unfair-commercial-practices-directive_en

Impact on freedom of expression

The evident vocation toward freedom of expression is many times a direct cause of governmental censorship forbidding social applications in some countries. So it happens that Twitter, Facebook, Instagram, YouTube or even some thematic websites are not allowed. Here apart from political, ethical, and philosophical issues may come to the fore the economic and financial aspect of entering that market adhering to the requested censorship or not¹¹.

Freedom of expression is usually associated with the terms hating, online libel, hoax, fake news this because the improper use of freedom of expression can generate such negative behaviours. Of course, such extensive and negative interpretation of freedom might generate some reactions that can be even worse than the problem itself. We must distinguish between two main branches “hating, online libel” and “hoax, fake news”, the first branch must be censored as it was at the time of Netiquette, the second, if not related to the first, is much more critical to be managed without the risk of infringing freedom of expression. A typical and sometimes concrete example is the establishment of a “commission” in charge for the fight against fake news, the one owning the “truth”, the risk in an “information society” is to cancel debates, silence alternate views and take a dangerous drift towards the “Pensée unique” or single thought.

Impact on businesses

The market is evolving in a very significant way. The diffusion of platforms if on one side creates new opportunities on the other side “kills” several existent businesses. The access to global service platforms creates a shortcut between offer and demand cutting out major part of the traditional added value chain, as it was long time ago because of malls it is now because of platforms. The big difference is that you don’t need to invest relevant capitals to feed your business, the key investment is the creation of the digital platform, the asset you own is the number of users both on the offer and demand side, this to do not consider the fiscal benefits they usually enjoy compared with the traditional retail system.

Following the schema of some of the recent revolutions the idea was: digital technology is disruptive cancelling several businesses, but new businesses will be created, the key point is that the specific nature of digital technology is actually creating less positions than the one eliminated. The visible effect now is an increasing number of workless people replaced by software and robots.

In addition, today digital tools are blurring the boundary between personal and professional lives, this effect is often termed “**time porosity**” or “spill over”.

Impact on commerce

An outcome of the merge of big data analytics and behavioural psychology is Internet of Behaviours (IoB). A very rough description of the IoB is the mash-up of three disciplines: Cyber Technology, Data Analytics, and Behavioural Psychology (Emotions, choices, augmentations, and companionship).

From a behavioural psychology standpoint, the IoB tries to comprehend the data acquired from users' online activities sometimes merged with IoT data. This mix offers important information on client behaviours, interests, and preferences Consumer data may be gathered from a range of sites and technologies, including a company's website, social media profiles, sensors, telematics, beacons, health monitors, and a variety of other devices.

¹¹ E.g., markets potentially offering “billions” of additional customers. Sometimes the censorship is not declared but the bandwidth devoted to the specific service or website is so narrow that it is practically impossible to connect.

Conclusions

In conclusion, don't you feel framed by such an "intelligent" environment? Social and communication media complete the panorama adding a "private depth" to the general fresco, ad-hoc defined tweets or posts may collect and analyse users' feedbacks to guide or anticipate citizens' actions and feelings. In recent times crowd data collection, open data, and big data, more or less anonymised, have provided the big framework was to collect all the different tiles. Online malls and delivery platforms offer, in addition, to analysing your browsing, the opportunity to save a "wish list" to better focus on the market trends. So, again don't you feel framed?