

AI for Good v/s AI for Bad

Alfredo M. Ronchi
Politecnico di Milano – Secretary EC MEDICI
alfredo.ronchi@polimi.it

Preamble

On this occasion my thoughts go to Evgeny Kuzmin the architect/creator of this relevant event. There are some key events that have characterised the recent period one of these is the so-called digital transformation considered the natural evolution of the current society in the light of a pervasive technology like digital technology. Within this transformation one of the most relevant potential impacts on society is probably due to Artificial Intelligence and related technologies.

Artificial Intelligence and Machine Learning

The “AI & ML” sector seems to have, currently, the most relevant impact on large part of society involving privacy, freedom, labour, security, lifestyle, and more. It is not under question the added value and the achievements due to digital technology; this contribution looks at the digital world from the humanities side.

Several services are already managed by artificial intelligence as well as decisions and even critical decisions are assigned to Artificial Intelligence, many times Narrow AI.

A.I. will increasingly pervade applications and devices. Few days ago, on June 2024, Sam Altman announced the integration of an Open AI chat GPT based AI assistant on iPhone. This trend will expand soon involving both everyday “tools” and high-end applications and services. If spell and grammar checkers have already created new dialects/languages, e.g. MS-English, AI is now generating one or may be multiple “creativities” accordingly with data that fed the system.

The infancy of the Big Brother

Artificial intelligence, expert systems and fuzzy logic were some of the keywords in the 1980s, at that time AI’s advances captured the interest of journalists being considered the seed of the “Big Brother” or the ignition of the progressive slavery of humanity ruled by machines. The

concern that time was “Artificial intelligence design will improve itself to think faster and deeper, then the improved version would improve itself, and so on, exponentially”.

One of the possible biases in this branch of computer science is due to the choice of the term “artificial intelligence” that ignited the understanding that there were two “intelligences” one human and another artificial potentially competing to rule the world, as some Sci Fi movies use to depict.

At the end of the 1970s early in the 1980s we started hearing about computers writing their code while running “intelligent” applications, basically auto-instructing themselves. It was the time of “Big Blue” a chess gaming expert system developed in the middle of the 1980s by Carnegie Mellon University running on a purpose-built IBM supercomputer. Early in the 1980s a typical computer science thesis was dealing with hypertext or chess gaming software. A relevant number of computer scientists, mainly coming from cybernetics, were experimenting new languages and new approaches to make machines “intelligent”, much more like humans. One of the first citizens’ “close encounters” with artificial intelligence dates to 1968, “2001: A space Odyssey” a science fiction movie directed by Stanley Kubrick, written by Kubrick and Arthur C. Clarke¹.

Later it was the time of Symbolics computers, LISP and PROLOG programming languages. Craig Reynolds, from the Symbolics Graphic Division, devised an algorithm that simulated the flocking behaviour of birds in flight. “Boids” made their first appearance at SIGGRAPH in the 1987 animated short “Stanley and Stella in: Breaking the Ice”².

Nowadays we talk with our “buddies” “*Hi Google: Set temperature to 24C*”, “*Hello Mercedes play Disco music*” or even closer to science fiction Alexa taking full control of our daily life. If we refer to more recent science fiction this “world of machines” might generate big concerns about the future, I don’t refer to “Terminators and Skynet by Cyberdyne” but simply remind the movie “WarGames”³ or “Eagle Eye”, in both cases artificial intelligence leads machines to apparently logical outcomes that are in direct contrast with human basic principles⁴.

AI dilemmas: ethical and fair AI

While AI will benefit citizens, businesses, and public interests it will create risks to fundamental rights. AI poses some concerns mainly regarding the field of ethics [1 – Stuckelberger 2018]. A short list of key dilemmas of AI ethics can start from AI bias (e.g. gender, culture, etc.), automated non supervised decisions, autonomous vehicles behaviours (land, air, water) in case of crisis and related responsibilities, impacts of general artificial intelligence, ethical data

1 This was probably the origin of the idea humans ruled by computers. The dialogue between HAL 9000 and the spaceship officer Dr. Dave Bowman, he used to interact with HAL by voice calling “Hello, HAL. Do you read me, HAL?”—HAL: “Affirmative, Dave. I read you.”, Dr. Dave Bowman: “Open the pod bay doors, HAL”, HAL: “I’m sorry, Dave. I’m afraid I can’t do that” . . . We all know what happened later.

2 Stanley and Stella: <https://youtu.be/3bTqWsVqyZE>, last accessed January 2024.

3 Science fiction movies already proposed similar scenarios e.g. Wargames (1983 American techno-thriller film directed by John Badham) or Eagle Eye (2008 American action-thriller film directed by D. J. Caruso)

4 We can recall the Isaac Asimov’s “Three Laws of Robotics”: 1) A robot may not injure a human being or, through inaction, allow a human being to come to harm. 2) A robot must obey orders given it by human beings except where such orders would conflict with the First Law. 3) A robot must protect its own existence if such protection does not conflict with the First or Second Law.

sourcing use, analytics and reuse, this to do not mention malicious use of AI to exploit resources, leverage on deep fakes and nudging, influence opinion dynamics and perform high-end social engineering.

More in detail, a potential ethical concern could be the idea to solve serious ethical dilemmas simply referring to an AI proxy to receive suggestions on how to behave and feel released from a personal ethical analysis and related responsibilities e.g. in the health sector.

Safety and security standards for such devices are not set actually; another typical ethical dilemma refers to how will two cars, both from the same builder or not, behave in case of imminent collision. The algorithm must decide which one of the two can be sacrificed the one with a baby or the other with a grandfather? which will be the decision of the algorithm and what about the implemented logic? There might be a “creative” solution due to human mind? How much technology and A.I. overlap moral and ethical aspects? This example depicts even the potentially different “logics” due to different cultural models (eastern / western).

More at practical level GPT can support businessman creating an almost perfect contract but this will be generated accordingly with the US regulations.

On AI for Good international event the UNESCO session dealing with this problem proposed to create several AI systems fed with different cultural models to ensure equal opportunities to different cultures. This aspect it is not far from the potential bias due to the “mainstream content” delivered online throughout the world via the Internet. Such training of young generations could have the potential side effect to jeopardise cultural identities. Here it comes the potential role of the Metaverse. Since more than two decades we have been wrapped in our personal cyber-sphere in a kind of symbiotic relation. One of the potential drawbacks due to lives spent in cyber-bubbles is the increasing cyber-mediation of human relations. Citizens experience the world, thanks to a cyber device mediated approach, the “new reality” is the one delivered by devices. Cyber-loneliness, one of the foreseeable risks is a kind of addiction to this “parallel life” training users to shift from real to Meta-life blurring the border between them. Metaverse can propose a “new normal” that once accepted in the Meta-life might be accepted in the Real life. The same of course is valid for information and opinion dynamics, especially if perceived as real and trustable.

Social media have drastically changed the way opinion dynamics evolve, they have become a battlefield on which opinions are, often violently, exchanged. The use, abuse or misuse of AI in social media can profoundly impact society. In the “new reality” there is a concrete and present risk to manipulate opinions thanks to digital media as well as to impact the decision-making process even through nudging. 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.

How to shape AI regulations

Recently it appeared that Microsoft basically acquired OpenAI, former OpenAI leaders Sam Altman and Greg Brockman are now working at MSFT. This acquisition multiplies the investments in such technologies extending their use in different scenarios. Open AI recently released a new AI tool that performs far better than previous ones, so developers switched to this product considering their own achievements already obsolete.

The extensive use of Artificial Intelligence, Machine Learning and Big Data, apart from several ethical issues, can lead to some relevant drawbacks if not carefully considered in the design and implementation phases. Today Artificial Intelligence is growing at a phenomenal speed. Mind boggling events like introduction of GPT-4⁵ represent unprecedented challenges. Hence, regulation of Artificial Intelligence including Generative AI becomes of crucial and critical importance for all stakeholders in AI ecosystem.

Some years ago, to promote awareness and stakeholders' cooperation the International Telecommunication Union (ITU) launched an initiative named "AI for Good". ITU organised several meetings and workshops aimed to discuss on specific uses of AI in different fields offering the opportunity to connect from remote and attend some live presentation of AI applications. Last May was held in Geneva the AI for Good general conference and exhibit, this event provided an interesting insight on the state of the art of this set of technologies applied in a wide range of sectors. One of the most appealing topics was the growing interest in AI regulations, this interest equally shared between citizens and AI developers.

On the regulatory side overall aim is to cover all AI, including traditional symbolic AI, Machine learning, as well as hybrid systems. In such specific field EU Regulations adopt a risk-based approach to regulation⁶. AI should be as neutral as possible to cover techniques that are not yet known/developed. We know that usually regulators do not intervene too early on new branches of technology to do not bias their natural evolution, this is true even in this sector.

Different international organisations and governments are working on regulations. More in detail UNESCO launched the AI Initiative and published different reports on these topics including UNESCO Generative AI, UNESCO Generative AI in Education, and more.

IEEE published Ethically Aligned Design, a vision for prioritizing human well-being with autonomous and intelligent systems. Similar initiatives were due to GPAI The Global Partnership on Artificial Intelligence, OECD Artificial Intelligence and Robotics, and UNICRI United Nations Interregional Crime and Justice Research publication "Toolkit for Responsible AI Innovation in Law Enforcement" or "AI for Safer Children".

⁵ <https://openai.com/index/gpt-4/>

⁶ <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

Lastly the European Commission was the first institution to invest in finding an agreement among different stakeholders to reach a first regulatory framework putting humans' rights and ethics as key reference. EU released the first edition of the EU AI ACT, the EU vision on AI can be summarised as “beyond making our lives easier, AI is helping us to solve some of the world's biggest challenges: from treating chronic diseases or reducing fatality rates in traffic accidents to fighting climate change or anticipating cybersecurity threats.”⁷

Global AI or Local AI

We feed ML systems mainly with big data from western countries this can lead, as happens in case of minoritized languages, to the disappearance of other “intelligences”, how to remove biases in machine learning models that could potentially discriminate against under-represented groups. Citizens are increasingly using AI “bots” to carry out different activities ranging from writing a poem to creating a deep fake. One of the potential scenarios in the near future is to face the geometric proliferation of document due to this ghost author. How can we identify a human “product” from a machine product? Is AI to be considered as a co-author? Lawyers are already animating the debate together with other stakeholders. “Local content” will be soon generated by “local” bots? These requests many times create potential conflicts on the side of IPR, do we need to include the BOT as an author, as it is?

Publishers and event organisers are asking the contributors to sign a declaration about the use or not use of AI based content (text, images, movies, etc), is this simply an integration of paternity (human + cyber), or is it a release of responsibility related to IPRs or other?

Some researchers suggest issuing a regulation to impose the insertion of an invisible watermarking in each AI generated output.

One of the recurrent topics on AI for Good conference was about AI for Good V/s AI for Bad – Malicious uses of AI [2 – Ronchi 2021], the fight between AI generated fakes and AI debunked ones? [3 - European Union 2016]. If AI can transform an engineer in a “super engineer” or a medical doctor in a “super medical doctor” we are ready to face “super criminals”.

This process takes us to consider the potential impact on ethics. The new ethics calls into question personal free will and freedom of choice; traditional cultural regulators of social relationships and processes are being replaced by automated social algorithms. Public perception is shaped more by addressing predetermined feelings and opinions rather than facts [4 – Makkuni 2018]. Furthermore, a massive decrease in the level of critical thinking and the emergence of waves of information epidemics are observed both nationally and globally.

⁷ Artificial Intelligence for Europe

Summarizing

This is not a complete overview on the key aspects and trends that appeared in recent times, off course taking into consideration each single technology and trend there are not specific concerns and technology seems simply to ease our daily life but getting much more in depth of each single innovation or putting together all the visible “tiles” of the “new normal” mosaic we can be concerned. If on one side the whole architecture is based on cyber tech, with all the potential risks it implies, on the other side cyber-world rules can express a power that no one of the “rules” in history had.

References

- 1 Christoph Stuckelberger, Pavan Duggal (2018), Cyber Ethics 4.0: Serving Humanity with Values, ISBN 978-88931-265-8, Globethicsnet
- 2 Ronchi A.M., (2021), Soft but still concerns, proceedings International Conference on ‘Homeland’ Security Emerging Trends, Challenging Aspects - Hasan Kalyoncu University, Turkey 2021 // Ronchi Alfredo M., 2022, From Ingsoc to Skynet it is not only science fiction: From novels and science fiction to quasi-reality, UNESCO IFAP Intergovernmental Council 2022
- 3 European Union (2016) Joint Framework on countering hybrid threats a European Union response, 2016
- 4 Ranjit Makkuni, Betrayed IT Revolution, <https://www.facebook.com/watch/?v=192288409296674>