Ethics and Artificial Intelligence

Alfredo M. Ronchi
Politecnico di Milano, Piazza Leonardo da Vinci 32, Milano, 20133, Italy
Tel: +39 393 0629373, Email: alfredo.ronchi@polimi.it

Abstract: Starting from the description of the two terms the document focusses on practical implementation of AI and related benefits and concerns. Concerns are mainly pertaining ethics, ranging between ethical dilemmas and different potential bias influencing AI outcomes. Key players approach to AI and the attempt to regulate this subject. The idea of a global AI system and the need to develop local AI systems.

Keywords: Ethics, Human Rights, Privacy, Artificial intelligence, Machine Learning, Digital Transformation.

Introduction

To contextualise the topics described in the next paragraphs it is proper to start from the two terms composing the title of the present paper: Ethics and Artificial Intelligence.
Ethics, in a very basic definition, is one of the branches of philosophy that concerns human behaviour. The term Ethics finds is root / etymology in the ancient Greek term "ἦθος," (ethos), this term refers to the characteristic spirit of a culture, era, or community as manifested in its beliefs and aspirations.
The term Ethics is often associated with the term Moral as in the case of Ethics as Moral Philosophy, the philosophical study of the concepts of moral right and wrong and moral good and bad. Ethics subject consists of the fundamental issues of practical decision making, and the ultimate value and the standards by which human actions can be judged right or wrong.
The term Moral is often associated with religions, cultures, professions, or virtually any other group that is characterized by its moral outlook (e.g. Medical Doctors, Biologists, Judges, Engineers). Philosophers developed their approach to ethics as part of their theory starting from the early stages of thinkers. At the Roman age we can mention the treaty "De finibus bonorum et malorum (*On the Supreme Good*)" [1 – Cicero 45 b.c.]. The work is structured in "books" based on the Aristotelian structure involving two orators, one sponsoring a thesis the other one refutes. In the third "book" Cicero let the orator Marcus Porcius Cato Uticensis (Cato the Younger 95 B.C., Rome–April 46 B.C., Utica) sponsors the Stoics' approach to good and evil. In the fourth "book" Cicero considers that this approach is not to be refuted but must be partially amended because it is too abstract and far from everyday reality. Stoics both do not consider pain as evil and assert that all the sins are equal. Stoics' ethics identify bliss in virtue, and this encompasses duties and sacrifice. Bliss is untroubledness, the waiver of the passions, not believing in needs, contempt of adversity, suffering and disease. This short excursus outlines the tight relation between ethics and groups or communities.
How it happened that ethics entered the cyber domain? Since the origin of computer science there was apparently no relation between this branch of human knowledge and philosophy, ethics and more in general humanities. Computer science seemed to do not impact on society at least not enough to generate interests or concerns by humanists, something happened at the birth of the World Wide Web. Since the early stages, the WWW attracted knowledgeable people from a wide range of cultural sectors. On the third WWW Conference (1993) held in Darmstadt (D) mayor part of participants come from humanities, they found the concept of

"Web" very appealing and useful in their specific domain of knowledge. At least since that time the humanism entered the world of computer science. A turning point was probably the World Summit on the Information Society held in Geneva ten years later in 2003 mainly thanks to the multistakeholder approach that involved in the process civil society and a large set of expertise including humanities, a specific interest was devoted to Cyber Ethics and to pursue a human centred approach [2 – Stuckelberger 2018].
Cyber Ethics become one of the relevant topics to be discussed and addressed, this led to the creation of a specific action line "AL10 Ethical dimensions the Information Society" that since that time, working in tight cooperation with other action lines, represents the humanism side of cyber technology. More recently thanks to the activity carried out both within the WSIS Forums and some research centres[1] the focus has been extended to the whole Digital Humanism.

Having introduced the first term, Ethics, we need to focus on the second one, Artificial Intelligence. Considering the ontology point of view Cyber Technology is a new entity, a new class of objects. In the last decade, philosophers and humanists started to professionally deal with computer scientists and innovators. These scholars usually considered the medium and long-term impacts of technologies on society. The emerging technological trend in autonomous vehicles, robots, machine learning and artificial intelligence may pose significant ethical problems to innovation.
After the previous interest in the 1980s, sometimes evoking the negative role of this branch of technology, the trend concerning the "reborn" AI and ML is generating concerns. If on one side AI will benefit citizens, business and public interests on the opposite side creates some risks for the safety of consumers and users, for fundamental rights, potentially releasing humans from ethical dilemmas. Some scholars suggest that the definition of AI should be as neutral as possible to cover techniques which are not yet known/developed. The overall aim is to cover all AI, including traditional symbolic AI, Machine learning, as well as hybrid systems.

How can be defined AI? On the legal side we can refer to the "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence"
"The term "artificial intelligence" or "AI" has the meaning set forth in 15 U.S.C. 9401(3): a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine- and human-based inputs to perceive real and virtual environments; abstract such perceptions into models through analysis in an automated manner; and use model inference to formulate options for information or action."
While "the term "AI model" means a component of an information system that implements AI technology and uses computational, statistical, or machine-learning techniques to produce outputs from a given set of inputs."

The European Commission provides a different definition of AI, EU Regulations adopt a risk-based approach to regulation. "'AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments;"

1 Vienna manifesto : **tps://caiml.org/dighum/dighum-manifesto/**

The European Commission clearly distinguish between "simpler traditional software systems or programming approaches ... that are based on the rules defined solely by natural persons to automatically execute operations. A key characteristic of AI systems is their capability to infer".

China is investing relevant resources to lead the AI domain. One of the definitions of AI due to Chinese researchers is "*Artificial intelligence (AI) aims to mimic human cognitive functions and execute intellectual activities like that performed by humans dealing with an uncertain environment.*[2]". The Chinese approach to this technology is in line with Chinese culture and foresees, in recent rules, that AI must always be under human control (since the advent of Generative AI as a precautionary approach). Positive duties for applications to promotes social cohesion, additional regulations prohibitions on algorithms that create division.

Benefits and drawbacks

No doubt that AI can benefit society performing an incredible number of duties including duties almost impossible to be performed by other means like big data analytics aimed to identify potential patterns, as it happens in health sector and biology, or decoding of ancient inscriptions in unknown languages. Concerns on AI start to focus if we consider potential 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 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. An additional aspect to be carefully considered is the quality and origin of data used to feed and train the system, these data will directly influence the outcome. Here it comes the role of supervised or not supervised systems and the opportunity to intervene as humans in the automatised decision making process. Automated decision systems based on AI are many times part of a network of sensors and actuators, the design of such solutions must carefully take into account malfunctions of one or more devices and behave properly (e.g. alerting humans and activating risk mitigating procedures).

Safety and security standards for such devices are not set actually; a typical ethical dilemma refers to how will two cars 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). Among the others, 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.
More at practical level GPT can support businessman creating an almost perfect contract but, due to its training, this will be generated accordingly with the US regulations.
Having timely identified the potential relevance of AI both at innovation and national soft power and security level, some country devoted a specific ministry to Artificial Intelligence[3].

2 Chen Shuang - School of Chinese Medicine, University of Hong Kong, Hong Kong 999077, China
3 The Vice President and Prime Minister of the UAE and Ruler of Dubai, Sheikh Mohammed bin Rashid Al Maktoum established the State Ministry of Artificial Intelligence and appointed as Minister Omar Sultan Al Olama. Under this Ministry United Arab Emirates are developing several initiatives to promote AI studies and expertise as "Learning Artificial intelligence" and "The National AI Strategy 2031"

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. 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.

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.

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 [3 – Ronchi 2021], the fight between AI generated fakes and AI debunked ones? [4 - 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”. Artificial Intelligence and Machine Learning are playing a relevant role even in this sector as AI v/s AI.

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 [5 – 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.
Recently it appeared that Microsoft basically acquihired 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 ML solution that performs far better than previous ones, so developers switched to this product considering their own achievements already obsolete.
Today Artificial Intelligence (AI) is growing at a phenomenal speed. Mind boggling events like introduction of CHAT GPT-4 represent unprecedented challenges. Hence, regulation of

Artificial Intelligence (AI) including Generative AI becomes of crucial and critical importance for all stakeholders in AI ecosystem.
As an additional case study let's consider "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. 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. Public perception is shaped more by addressing predetermined feelings and opinions rather than facts.

Closing remarks

Finally, both industry experts and government bodies are calling for specific regulations and laws to be enacted to ensure that AI-assisted systems deliver positive and ethical results., one of the most relevant voices in this field was due to San Altman "OpenAI's Sam Altman Urges A.I. Regulation in Senate Hearing" [May 2023] or a newspaper level "The world wants to regulate AI, but does not quite know how" [The Economist - Oct 24th, 2023].
Post-reality is changing the system of values with the "new normal" (semantic shifts, etc.), that is, by definition, politically correct. New ethics is putting personal free will and freedom of choice under question; traditional cultural regulators of social relations and processes being displaced by automated social algorithms (increasing role of algorithms and ML). Widespread simplified virtual mock-ups and simulacra are not only blurring the borders between the real and the digital world but also led to a mass collection of data for managing people's behaviour (evaporation of privacy, data protection). It has also been contributing to the increasing level of conflict in both the society (between individuals and groups – haters, discrimination) and among states, soft concerns and more.

References

[1.] Cicero Marcus Tullius, (45 B.C.), De Finibus bonorum et malorum. LOEB Classical Library. https://archive.org/ details/de nibusbonoru02cicegoog
[2.] Stuckelberger Christoph, Duggal Pavan (2018), Cyber Ethics 4.0: Serving Humanity with Values, ISBN 978-88931- 265-8, Globethics.net
[3.] 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
[4.] European Union (2016) Joint Framework on countering hybrid threats a European Union response, 2016
[5.] Ranjit Makkuni, Betrayed IT Revolution, https://www.facebook.com/watch/?v= 192288409296674