AI, trust and communication

Alfredo M. Ronchi
Politecnico di Milano, Piazza Leonardo da Vinci 32, Milano, 20133, Italy
Tel: +39 393 0629373, Email: alfredo.ronchi@polimi.it , ORCID 0000-0003-1230-4338

Abstract: We can start from a general approach to the concept of thrust and specifically trust in algorithms, "intelligent" automated procedures to reach neural networks black boxes and ML in its various declinations. Till what extent intelligent systems can positively solve concrete case or "organisms" based on sensors and AI are fail proof. It is wise to exclude from decision processes humans? AI as the essential keyword to ensure market success. AI considered a potential cyber-Leviathan dominating humanity thanks to its high "intelligence far exceeding any human one"? The term artificial intelligence generates the idea that this technology will compete and exceed the human one thanks to the ability to self-improve even recursively. How currently is defined AI in different countries. Narrow AI and General AI. Can digital tools express creativity? Can AI be considered an author? This takes us to consider information, freedom of expression, and misinformation including fake news and deep fakes. Opinion manipulation and nudging complete the panorama.

Keywords: Digital Transformation, Artificial Intelligence, Ethics, Human rights, Freedom of speech

From Star-wars to Quantum Computing and AI

Once upon a time it was the race to space, to the moon, to technological innovation, to super computers, more recently star-wars, nowadays it seems that the competition among key players is in the domain of quantum systems and artificial intelligence. Deep Seek and Open AI, Google, META & friends are investing in LLMs like Chat GPT, Mistral, LLAMA, and more. Hardware and microchip developers once named Symbolics machines now are represented by neural processing units (NPU) like NVIDIA or AMD. On September 2025 NVIDIA and INTEL announced, "*a collaboration to jointly develop multiple generations of custom data center and PC products that accelerate applications and workloads across hyperscale, enterprise and consumer markets*." Few days later OpenAI and NVIDIA announced a strategic partnership to deploy 10 Gigawatts of NVIDIA Systems. Investments and joint ventures depict a potential impressive development of this sector and related impacts on society.

The origin – it was a game changer

Following the studies on Turing[1] machine in 1950 a first release of the Touring Test was published by A. M. Turing on Mind[2], the incipit of the article is "*I propose to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. ...*". For the first time the idea to break the rule of rigid procedures and algorithms unleashing the hidden power to identify "invisible" patterns and create associations of items. Since that time: expert systems, fuzzy logic, chess challenge between Garry Kasparov and Deep Blue[3]. The initial understanding of AI was "the ability of computers to dynamically

[1] Alan Mathison Turing (1912-1950), British mathematician Alan Turing, c. 1930s. Turing did the earliest work on AI, and he introduced many of the central concepts of AI in a report entitled "Intelligent Machinery" (1948)

[2] Alan M. Turing, Computing machinery and intelligence, Mind, 59 October, pp. 433-460, 1950. https://academic.oup.com/mind/article/LIX/236/433/986238

[3] https://www.ibm.com/history/deep-blue

instruct themselves creating code". A.I. advances captured the interest of journalists being considered the seed of the "Big Brother" or the ignition of the progressive slavery of men ruled by machines. This vision generated the concern "General artificial intelligence design and generative AI will improve itself to overcame human intelligence[4]?" Are we going to face what several times was depicted in the science fiction movies?
We don't know if the concerns often related to AI, considered a potential cyber-Leviathan dominating humanity thanks to its high "intelligence far exceeding any human one", is due to the term that was associated long time ago to the early studies and applications in this field. The term "artificial intelligence" probably ignited the understanding that there were two "intelligences" one human and anther artificial potentially competing to "rule" the world, as some Sci Fi movies use to depict like "*2001: A space Odyssey*" due to HAL[5], War Games, Eagle Eye or even Lucy[6].

AI actually

The added value and results achieved thanks to artificial intelligence are not in question, just think about the use of artificial intelligence to identify hidden patterns in DNA or collect and summarize a huge number of documents and much more; we will look at artificial intelligence from a humanistic perspective, considering the medium and long-term impacts on society and the potential disadvantages or unwanted side effects[7] with reference to human rights, security, ethics, social, cultural, economic and environmental aspects. Is or will the sudden evolution or revolution of the cyber world, now driven by artificial intelligence and related technologies, be fully beneficial to humanity, or are there unforeseen and unwanted side effects?
AI and ML can offer relevant benefits to humanity but, after a first phase of free evolution and identification of positive evolutionary trends, they must be carefully framed in their potentiality and regulated. This regulatory action will address both development and deployment of their applications to avoid potential drawbacks. As an example, citizens consider Generative AI able to express "creativity" how can we ensure that its "creation" comply with ethical principles and will not generate harm?

A marketing passepartout

AI as the essential keyword to ensure market success. The "commercial" power of "AI based" tools and equipment is evident. Thanks to the blurred border line between programmed logic behaviour and artificial intelligence a relevant part of products and services are nowadays "AI based". Consequently, becomes important to set some guidelines framing this sector. 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[8]"

---

[4] 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)
[5] A Space Odyssey is a 1968 epic science fiction film produced and directed by Stanley Kubrick based on an Arthur C. Clark novel – HAL acronym is formed by the chars preceding IBM - The spaceship officer Dr. Dave Bowman 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.
[6] 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), Lucy (2014 French Sci Fi movie written and directed by Luc Besson)
[7] Ronchi Alfredo M., (2021) – "From Ingsoc to Skynet it is not only science fiction: From novels and science fiction to quasi-reality" – Proceedings UNESCO IFAP conference in Khanthi Mansjisk
[8] https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/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[9] . “‘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.”
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.[10]”. 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.

To conclude let’s consider the standpoint of Saudi Arabia that was one of the first country to establish the National Centre for AI (NCAI). What is AI? “*AI is systems that use technologies capable of making predictions, generating content, providing recommendations, or making decisions with different levels of autonomy*.”

AI daily life invasion

Artificial Intelligence will increasingly pervade applications and devices. On 18 May 2023 Open Ai released a ChatGPT APP on IOS, on June 2024, Sam Altman announced the integration of an Open AI Chat GPT based AI assistant on iPhone, in 2025 META launched AI on WhatsApp providing major part of the typical AI services, above we already considered some of the following steps and developments. AI extensions are offered as a kind of bonus or add-on to any product from fridges and washing machines to internet service provision. Automotive industries are starting to extend the use of AI out of autonomous guide. Big tech companies like Google and Microsoft are upping their game. Anthropic[11], Mistral[12], and OpenAI are continuing their winning streaks.
This trend will expand soon involving both everyday “tools” and high-end applications and services. Several services are already managed by artificial intelligence as well as decisions

[9] Source: Artificial Intelligence Act https://artificialintelligenceact.eu/article/3/
[10] Chen Shuang - School of Chinese Medicine, University of Hong Kong, Hong Kong 999077, China
[11] https://www.anthropic.com
[12] https://mistral.ai

and even critical decisions are assigned to Artificial Intelligence, many times based on Narrow AI[13]. Large Language Models[14] are quickly developing offering a wide range of "services" to even not skilled citizens. These "end-users" will accept the outcome of AI mainly as an "oracle" with a limited or absent ability to analyse or criticise it. Apart from potential conflicts with ethics, errors and drawbacks a potential mid-term effect should be to lose the ability to perform similar tasks traditionally having delegated these tasks to AI.

Artificial Intelligence v/s Human Intelligence is this the challenge?

The recent polarisation between the supremacy of human intelligence and AI overtaking human intelligence take us to consider some of the evolutionary phases of AI from Narrow AI (NAI), Expert Systems, Machine Learning, Deep Learning, LLM, General AI, and Generative AI all these steps, more or less in this order and terms extended the abilities of these systems from rigid almost algorithmic behaviours to a kind of "freedom" demonstrated by LLMs, general AI and more evidently by Generative AI. Scientist and Researchers foresee two more frontiers to be reached the "Human Brain" functionality model aimed to recreate an even growing web of synapsis and the ultimate level, "Self-Consciousness".
This latter level of evolution is probably to one freeing the machinery from any dependence from humans and its "architect". One of the most well-known examples of self-consciousness is probably represented by the scene in Stanley Kubrick's "2001 A Space Odyssey", captain Dave Bowman starts to disconnect HAL 9000 computer removing the computer's boards. HAL 9000 proves to have self-consciousness understanding that he is going to "die". "My mind is going... I can feel it", "I'm afraid. I'm afraid, Dave. Dave, my mind is going. I can feel it. I can feel it. I can feel it. I'm a... fraid". This theoretically foreseeable step is, as the previous one, Human Brain, not yet included in near future reachable objectives. These steps were the turning point, in several Sci-Fi movies, to enable the rebellion of machineries against humans. The debate is still open and polarized, with some industry leaders clearly calling for AI regulation, especially to avoid a critical gap between predictable human behaviour and AI hallucinations.

Another of the most debated concerns is the one concerning potential biases like gender, race, cultural identity and privacy. AI systems start their processes from the huge dataset that was made accessible to feed the system. All these systems base their activity on data, often big datasets, to be analysed as originators of the "outcome" of the system. The quality and characteristics of the dataset is one of the key factors to eliminate or at least minimize biases and increase trust. Additional potential causes of biases and deviations can be rooted in the algorithms performing the process.
If we refer to AI systems as a black box that processes input data analysing a huge data set to provide an outcome that it is not supported by process explanation, how can we be sure that no biases were involved in the process? How can we investigate if any potential bias influenced the outcome of the system?
Of course here we need to refer to the different learning models utilised in AI: in unsupervised learning the system is fed with data and finds patterns and predictions without any supervision, supervised learning uses labelled datasets effectively "supervising" the process with correct input-output pairs, and in reinforcement learning an "agent" learns to make sequential decisions in an environment by trial and error to maximize a cumulative reward signal.

---

[13] Narrow AI and General AI are one of the typical differences in AI domain – the first has a in depth but limited knowledge.

[14] LLM are advanced artificial intelligence (AI) programs trained on massive datasets to understand, generate, and process human language and other content.

On the citizens' side the evolution of similar problems may change through time, at the beginning citizens will assign a high level of trustworthy to the system but if some evidence distrusts the system citizens will start to ask for explainability of the results. Explainability was one of the key requisites in the 1980s artificial intelligence early applications, the quest for explainability is back in research labs today. The aim is to open the black box or make it transparent to end users.

Furthermore, artificial intelligence, among its many uses, often supports decision-makers; decisions are typically influenced by our culture and traditions. Once cleared from any actual or potential bias our AI platform how to ensure "AI cultural identity"? Nowadays major AI platforms are fed with western data; this can generate the concern that other cultural models are not adequately represented in our AI platforms. Do we need to create several different AI platforms representing several other cultural models? In 2023 Jensen Huang, boss of Nvidia, proposed that every country should have its own artificial-intelligence system, trained on domestic data, aligned with national values and culture, built and hosted by local infrastructure. A similar suggestion was supported by some UNESCO delegates on the WSIS UNESCO panel in 2024 and replicated in some of the AI for Good sessions.

Creativity. Can machine be creative?

Since a long time, artists and scientists invested time and resources in developing generative AI to create artefacts, others merged AI with artificial life to develop self-evolving "creatures" or self-behavioural digital actors. Typical examples are the research developed long time ago by Demetri Terzopoulos[15] in the field of AI and artificial life and the movie "Star and Stella[16]" (1987) by Symbolics the first example of animation movie created using digital actors with autonomous behaviour far before "Titanic" top deck promenade and the "Lord of the Rings" Uruk Hai army.

Artificial intelligence can boost our creativity in a broad sense, can perform different added value activities in different fields as an example: performing arts, interpretation of huge datasets pertaining artefacts to extract invisible patterns and generate new artefacts, a wide set of autonomous systems from hotels' guest assistants to worms cleaning pipes there are no specific limits to creativity and imagination quoting Albert Einstein "Logic will take you from A to B. Imagination will take you anywhere."

Again, even in the field of creativity we need to ensure "different creativities" accordingly with cultural identities. These differences are often evident in international contests where cultural background, sensibility and taste are clearly expressed even if the tools used by authors are exactly the same.

AI Ghost Author and IPR

Few years ago, after the experiences carried out by some researchers and artists in the field of generative AI a software tool disclosed to citizens a concrete and easy to use application of artificial intelligence, the chatbot ChatGPT. This application raised an immediate success and the community of users grown up exponentially, citizens were captured by a technology that before was considered cryptical, unreachable, limited to the restricted group of researchers and nerds. The application of ChatGPT was extended to various fields starting from the creation of homework, poems, contracts and more.

---

[15] https://scholar.google.com/citations?user=pKuBFaQAAAAJ&hl=en // self-animated pedestrians https://youtu.be/cqG7ADSvQ5o

[16] https://archive.org/details/stanleystellainbreakingtheice1987

Dealing with AI and content creation one of the potential scenarios is to face the exponential proliferation of documents due to this "ghost author", this will lead to new AI documents based on existent AI documents and so on diverging probably-more from human creativity.
As already pointed out nowadays the dataset training the systems is mainly western American so the outcome generated by these systems will pertain to that cultural environment.

How can we identify a human "product" from a machine product? Is AI to be considered as a co-author? 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? "Local content" will be soon generated by "local" BOTs?
In the meantime, 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 right (human + cyber), or is it a release of responsibility related to IPRs or other? Some researchers suggest governments issuing a regulation to impose the insertion of an invisible watermarking in each AI generated output. Policymakers around the globe are therefore pondering how to design and implement watermarking techniques to ensure a trustworthy AI environment. China has already taken steps to ban AI-generated images without watermarks; United States already enforced this option at least on specific events to be protected by AI generated bias (as in case of Political elections) developing effective labelling and content provenance mechanisms so that citizens are able to identify human or AI generated content.
A similar request to develop and deploy reliable content authentication and provenance mechanisms, such as watermarking, was requested by G7. The EU Artificial Intelligence Act (OJ 12 July 2024) requires providers and users of AI systems to enable the detection and tracking of AI-generated content[17]. As it happened in several fields now the access threshold to digital tools is lowered, much more people can perform similarly and, in short term the gap between the one that know-how and "the one that don't know what is shared" will be relevant so the risk to do not debunk fakes will increase. Once again there is a strong need to keep humans in the centre of the loop and respect human rights and ethics. As already outlined, nowadays, in the light of actual trends moral philosophy / ethics must be the always the key reference point in developing, customising and using technologies. Having considered AI, trust, creativity and IPRs let's move to information, freedom of expression, and disinformation.

Information, freedom of expression, disinformation

The incredibly rapid diffusion and massive use of the Internet and related technologies if on one side offered to citizens the unprecedented opportunity to reach broader audiences unthinkable before, on the flip side generated the "information flood" mainly due to newcomers in the field of information broadcasting eager to take advantage as much as possible from this opportunity, self-made journalists often having no idea about professional rules and ethical principles. Some of them considered this "freedom of expression", well-known as one of the key human rights, the "fifteen minutes of fame" promised by Andy Warhol. They do not use to follow the basic principles of journalism regarding information sources, accountability and thrust, echo chambers and "creativity" do the rest.
This "freedom" is often negatively associated with the terms hating, online libel, hoax, fake news especially after the diffusion of AI applications. This negative bias because the improper use of freedom of expression can generate action and behaviours aimed to regulate "freedom".

[17] https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS_BRI(2023)757583_EN.pdf

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 often connected with the polarisation between mainstream information and counter information or the so-called misinformation, is much more critical to be managed without the risk of infringing freedom of expression. The two opposite sides do not imply that one of the two represents the truth. As already expressed in other publications, we observe the massive decrease in the level of critical thinking and the emergence of waves of information epidemics, both at the national and global level (mainstream communication, limited contraposition, censorship, and fake news). These events are many times due to a well-defined and deployed manipulation based on AI applications.

Minitrue

This situation generated the idea to establish a kind of George Orwell’s Minitrue – Ministry of Truth in charge to identify proper information and apply censorship on the others. This trend, often due to the will to fight against fake news, was embraced mainly by private platforms that activated “AI supervisors/censor” automatically deleting any kind of content that, accordingly with the “owners’ vision” could be classified as improper. Public actions of censorship were mainly performed against entire platforms or services both to avoid external influences on citizens and guarantee technological independence [1 – Ronchi, 2019D]. These measures take us to focus on the rationale beyond filtering and potential infringement of human rights, freedom of information, opinion formation, and freedom of speech.
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” that become a “dogma”, 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. This action of censorship is mainly operated by AI filters. Furthermore, as we will outline in some following paragraphs, nudging enforced by AI and manipulation of opinion dynamics can led to the elimination of alternate stand points and the definition of “dogmas”. Truth that cannot be confuted, any attempt to propose a different approach is judged as an “heresy”. This generates a strong polarisation between good and bad, friends and enemies. The private ownership of key platforms and the right to manage privately any content and message, can drastically limit freedom of speech and other human rights. Some institutions and experts in human rights highlight the need to create independent state-own platforms ensuring freedom of speech and thoughts thanks to well-structured fully agreed rules and guarantees. As it already happens, mainly in case of private platforms, content is analysed, often even if privacy is ensured, by AI agents for different purposes: stats, profiling, marketing, and more. In this regards there are some initiatives, sometimes motivated by the fight against crime, to filter any digital communication before the encryption anyway evaporating privacy.

Opinion formation and Nudging

Opinion formation [2 – Peralta, 2022] evolved through time from the confrontation of opinion in open spaces where small groups of citizens use to debate about key topics and direct human interaction was the mean, to online challenges. Today opinion formation is a complex and dynamic process, often deeply influenced by mass media or the so-called mainstream information mediated by interactions among individuals in social networks, both offline and online. Citizens can find a proper echo-chamber to reinforce the endorsement of specific

opinions, may times be diverging from reality because of different biases introduced by contributors (e.g. extremist, trolls).
Opinion dynamics research [3 – Martins, 2009] has developed models of opinion formation, adjustment, and exchange for over sixty years[18]. The main mechanisms in these models are positive social feedback, opinion, and status homophily, negative social feedback, and the structure of the social network[19]. These models share the property that opinions are modelled as binary variables, finite sets, or low-dimensional vectors. While this is sufficient for toy models and allows to reproduce interesting qualitative behaviours, such as opinion cascades, opinion bottlenecks, gatekeepers, and opinion leaders, it is too coarse a description to connect these models to data from social media. Replacing these extremely coarse opinion representations by the devices that computer scientists use to represent opinion anyhow: knowledge graphs[20] it will be easier to analyse opinion dynamics.
Online information channels and 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. This environment is a breeding ground for misinformation, fake news, and trolls. The latter, which typically populate forums and thematic groups, are often driven by boredom, a need for attention, or even a sadistic desire to inflict pain [4 – Ronchi, 2019]. Trolls aim to provoke emotional reactions or divert the topic of a discussion from its original purpose. They can cause stress, contempt, and even cyberbullying and other forms of violence. Some researchers study opinion dynamics in social networks by large-scale analysis of social media using state-of-the-art methods of natural language processing and semantic artificial intelligence.

As an additional case study let's consider "nudging". The concept of nudge is intertwined with social psychology, consumer behaviour, behaviour economy and decision making. It can very basically be summarised as the art to address decision or behaviours seamlessly without infringing the right to freely choose. This concept born and exploited in the field of behavioural economy[21] was applied to several domains, the masterminds in this domain are usually termed "choice architects".
Choice Architects developed different techniques to guide citizens toward the foreseeable choice. We can start observing that human's behaviour it is not always fully rational, sometimes they do something that is not in their best interest, this discrepancy is termed "value-action gap". Daniel Kahneman [5 – Kahneman, 2011] identifies two different approaches to process information that lead humans to take decisions against their own interests: A) automatic, fast and potentially subject to environmental influences; B) reflective, slow, considering influences and outcomes.
In case of complex decisions overwhelming single human cognitive capacity, approach A is the choice. This choice can override explicit values and goals. This automatic behaviour is hard to be modified is do not intervene something that trigger a new behaviour. Consequently, heuristic judging techniques are based to judgmental heuristic; they use to modify the environment so

18 Banisch, S., Olbrich, E., Tat Dat, T., Yamshchikov, I.P., and Rezagholi, S. (2017). Opinion dynamics on abstract conceptual spaces. Report on the horizon2020 project ODYCCEUS (grant no. 732942).
19 DeGroot, M.H. (1974). Reaching a consensus. Journal of the American Statistical Association 69(345): 118-121. / Holley, Richard A., and Liggett, Thomas M. (1975). Ergodic theorems for weakly interacting infinite systems and the voter model. The Annals of Probability 3(4): 643–663 / Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. The Journal of Conflict Resolution 41(2): 203-226.
20 Ehrlicher, L., Wöß, W. (2016). Towards a definition of knowledge graphs. Proceedings of SEMANTiCS 2016.
21 From the 1990s several researchers pertaining to different domains from cybernetics to psychology - Thaler Richard, 2008, Nudge: Improving Decisions About Health, Wealth, and Happiness,

that when heuristic or approach A is used the choice automatically comply with the desired one. Typical actions to modify the environment are to set a default option or to comply with the social-proof heuristic[22]. Additional nudging techniques are to impact on visual perception making some options more visible or known. Some examples of nudging in our daily life are the arrangement of products in supermarkets and malls, the choice of cookies or licence approval regarding online products, the comparison of two sizes of the same product and related prices, to do not mention health, fitness, investments, advertisement, information and the already mentioned opinion dynamics. Nudging is since a long time used in the field of politics and consensus building, many times generating polarisation or even "dogmas", this is typically well conceived to be seamlessly accepted as the truth and joined "by definition" by citizens.
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.

Furthermore, the rapid implementation of AI poses potential ethics [6 – Stuckelberger, 2018] concerns, apart from potential bias, deep fakes and more, will citizens, for instance, solve serious ethical dilemmas simply referring to an AI proxy asking how to behave and feel released from a personal ethical analysis and related responsibilities?
Furthermore, more recently we started to discuss about the Global Digital Compact, this was one of the key topics on the WSIS Forum 2023 together with AI tools and their developments. "The Global Digital Compact that would set out principles, objectives and actions for advancing an open, free, secure and human-centred digital future, one that is anchored in universal human rights and that enables the attainment of the Sustainable Development Goals." The aim of the debate is to shape a shared vision on digital cooperation by providing an inclusive global framework for a sustainable digital future. We hope that the outcome of this debate will fully represent what is expressed in the statement above. The real impact of some of the emerging technologies such as machine learning and artificial intelligence are often emphasized by media, we must wait a little bit longer to assess the real impact of such technologies on society in specific fields such as employment or professional profiles ¡7 – Egger, 1996].
The challenges for the upcoming years are the ways to sustain the human's role and the inviolable right to freedom and personal privacy in an era of unlimited information gathering. Once again, the need to find a proper balance between humanities and technologies is omnipresent. Social sciences and humanities must establish a tight cooperation in designing or co-creation of cyber technologies always keeping humans in the focus.

Conclusions

A comprehensive and human-centred approach is essential for making digital transformation the concrete opportunity deliver high-quality public services like healthcare, education, and empower citizens to participate fully in the twin green and digital transition. More in general there is a need to reshape our technological landscape putting again humans in the focus. This is relevant in industrialised countries that already entered a technology centred innovation but even more relevant in countries that are now facing the digital transition process.
To mitigate risks such as monopolization of the information space, increasing levels of fake news and disinformation, strong power of online platforms, cyber threats and crimes, privacy breaches, strong

[22] Social Proof Heuristic is our tendency to copy the behaviour of others in uncertain or unfamiliar situations, it looks similar to bandwagon effect.

market disparities, a robust framework grounded in ethical principles is imperative. The described methodology adopts a horizontal and holistic approach that combines ethical considerations with extensive stakeholder engagement to develop a collaborative framework centred on transparency and human well-being. By disseminating findings widely together with media partners, the aim is to foster a larger shared understanding of digital humanism and encourage the integration of ethical principles into digital policies and regulations than today, that will empower individuals and organisations to shape a more equitable, trustworthy and human-centred digital future. By combining user-centric analytics and design, broad stakeholder involvement, and ethical considerations, this methodology aims to identify digital solutions that not only meet user needs but also contribute to a more equitable and sustainable future.

References

[1.] Ronchi A.M. (2019 D) e-Democracy: Toward a New Model of (Inter)active Society, ISBN 978-3-030-01595-4, Springer (D)
[2.] Peralta Antonio, et Al. (2022), Opinion dynamics in social networks: From models to data, Handbook of Computational Social Science
[3.] Martins André C.R. et Al., (2009) An opinion dynamics model for the diffusion of innovations, Physica A 388 (2009) 3225–3232
[4.] Ronchi A.M., (2019), e-Citizens: Toward a New Model of (Inter)active Citizenry, ISBN 978-3-030-00746-1, Springer (D)
[5.] Kahneman Daniel (2011) Thinking, Fast and Slow, ISBN 9781846140556, Allen Lane London
[6.] Stuckelberger Christoph, Duggal Pavan (2018), Cyber Ethics 4.0: Serving Humanity with Values, ISBN 978-88931-265-8, Globethics.net
[7.] Egger O. at al. (1996) Internet behaviour and addiction, Swiss Federal Institute of Technology