Governance and Ethics principles in the Cyber Domain

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

Preamble: It is not under question the added value and the achievements due to cyber technology (societal, intellectual, etc.); we look at cyber technology from the humanities side fostering a Digital Humanism approach. Posing our focus on processes that have led to governance agreements. Starting from internet governance, ongoing digital transformation to reach AI Governance Ethics and a selection of experiences carried out by international organisations, nations and single states.

Digital Evolution or Revolution?

As Klaus Schwab wrote in the preface of his book "*Shaping the Fourth Industrial Revolution*", "*The world is at a crossroads* ... "[1 - Schwab 2016]. We are facing a significant turning point, the convergence of cyber, nano, and biotechnology, coupled with the influence of "*Tangible and intangible impact of information and communication in the digital age*" [2 – UNESCO-IFAP] has fundamentally transformed society. Posing the focus on "cyber" that seems to have, the most relevant impact on large part of society involving privacy, freedom, labour, security, behaviour, and more.

Digital transformation is considered the natural evolution of the current society in the light of a pervasive technology like digital technology. Digital technology is intertwined with almost all the life sectors. Since its dawn the number of application and solutions based on such technology had a surprising rate of growth. The extended use of computers overlapped more and more any activity generating an impact on society.

In the last decades, social media and global content providers are "training" young generations offering all-over the world "global homogenised" content that will impact future generations and jeopardise cultural diversities. We are increasingly leaving the analog, face to face, paper-based world to enter the intangible digital mediated one.

Since more than two decades we are wrapped in our personal cyber-sphere in a kind of symbiotic relation. We experience the world through the mediation of cyber devices; the "new reality" is the one provided by them. By leveraging on laziness and relaxation citizens shop online, purchase food and drinks delivered on their table, "meet" friends on Zoom, interact with the "external environment" though social media and video clips.

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. Meta-life may propose a new normal that once accepted in the Meta-life could be accepted in the Real-life.

Social media have drastically changed the way opinion dynamics evolve. Opinion formation is a complex and dynamic process mediated by interactions among individuals both offline and online. Social media has become a battlefield on which opinions are exchanged, often violently.

A relevant part of digital transformation relies on platforms and standards [3 – Ronchi 2020] and it is intertwined with the "owners" of such platforms and standards. This can be considered a kind of monopoly not yet regulated a "grey zone". Platforms are mainly private, and the key ones

are concentrated in few countries creating a kind of “oligarchy”. The “control buttons” of our daily life are often outside the control of our nation state. So, in the digital transition, despite antitrust laws, there is a potential risk to fall under the control of few “private” key players.
In the “analogue” world we had different pipelines and “channels” to perform, thanks to different tools and means, our activities, in the cyber world the whole activity depends on a single “bottleneck”: cyber technology. This pillar is quite fragile and subject to attacks or suitable for top-down discrimination to selectively switch off the service. If this pillar will fail, malfunction or be switched off our life will suffer sometimes unpredictable problems, no cyber and consequently no digital identity, no bank account and social security, no service delivery, no news, and connection with other “entities”. A plan B in such a situation, if not present, will require long time to be implemented.

What about industrial machinery today fully computerised, or critical infrastructure management? In a cyber warfare scenario, it might be enough to dispatch on the network a tag like “sunrise” to collapse the whole target infrastructure[1]. Recently on the World Economic Forum 2024, Jeremy Jurgens [2], foresaw the possibility a global cyber-attack that will take us back to the stone age. Of course, even this message might be a fake news. Having briefly outlined the role and weight of Cyber Techs in our lives it is evident the need to clearly identify its governance.

Cyber world governance

The incredibly pace of growth of the Internet in the nineties and its potential impact on society posed the problem to clearly identify who was going to “rule the Internet”, or more precisely deal with the “Internet governance”. Conceived in Tunis in 2002 and implemented in two phases in 2003 Geneva and 2005 Tunis the World Summit on the Information Society (WSIS) has tried to find a proper solution.
One of the key aspects to be considered in similar situations is identify a sounding model for Governance, on the 2005 phase a specific organism was identified, a multistakeholder based forum, the Internet Governance Forum (IGF) together with the Multistakeholder Advisory Group (MAG). Since the first phase of the WSIS and similarly for the IGF, 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.
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.

Digital Humanism

Popper's seminal work in the late 1960s, which challenged the notion of technological determinism and posited humans as active architects rather than passive recipients of technology, has laid the groundwork for the prominence of digital humanism[3]. This perspective has shaped global discourse around information and communication technologies (ICTs) as exemplified at the World Summit on the Information Society (WSIS) that adopted an inclusive

[1] This to do not mention Wanna Cry and the registered domain iuqerfsodp9ifjaposdfjhgosurijfaewrwer gwea.com

[2] "Geopolitical instability makes a catastrophic cyber event likely in the next two years": WEF Managing Director Jeremy Jurgens on the Global Cybersecurity Outlook Report 2023.

[3] *Ronchi A.M. (2019). e-Services: Toward a New Model of (Inter)active Community, ISBN 978-3-030-01842-9, Springer (D)*

multistakeholder approach in establishing international cooperation and dialogue on ICTs, involved politicians, companies, and civil societies, emerging as a pivotal platform. Almost two decades later, it led to the formation of the Vienna Manifesto in 2019[4], which acknowledged the potential benefits of increased technological progress, information access, and societal transformation, highlighting the accompanying risks, such as the erosion of human rights through surveillance, monopolization, and top-down control, job displacement potential, privacy breaches, algorithmic bias, and the unchecked influence of big tech.

Notably, rapid advancements in AI and autonomous systems have precipitated a complex landscape where technological potential is counterbalanced by significant ethical and societal challenges. While these technologies hold immense promise for societal transformation, they also pose risks such as power concentration, disinformation, privacy breaches, and environmental degradation.

Similar concerns are expressed in the book "The Age of Surveillance Capitalism" by Shoshana Zuboff[5]. She highlights the drawbacks of monopolising information spaces, the concentration of power among online platforms, the escalation of cybersecurity threats, digital fragility, privacy infringements, market disparities, and the commodification of human experience for commercial exploitation.

The European Commission recently launched some calls addressing Digital Humanism including concerns related to the generalised use of robots[6] and humans robotisation[7]. This initiative relates to another EU initiative “Europe’s Digital Decade: digital targets for 2030”.

The policy programme Europe’s digital Decade sets concrete targets and objectives for 2030, to guide Europe’s digital transformation. It will act on four key sectors: skills, government, infrastructures, and business.
Skills: 20 million ICT specialists ensuring gender convergence, minimum 80% of population enjoying basic digital skills:
Business: digital transformation of businesses foresees the take up of technologies, 75% of EU companies using Cloud, Big Data, Artificial Intelligence. Referring to innovators grow scale-ups & finance to double EU Unicorns. To complete the foreseen scenario more than 90% of SMEs will reach at least a basic level of digital intensity[8].
Infrastructures: EU Digital Decades considers secure and sustainable digital infrastructures. The key parameters chosen by EU to evaluate digital infrastructures are *Connectivity:* Gigabit for everyone, *Computing:* first computer with quantum acceleration, *Cutting edge Semiconductors:* double EU share in global production, *Data - Edge & Cloud:* 10,000 climate-neutral highly secure edge nodes.
Government: full digitisation of public services, all citizens having access to their medical records online, all citizens having access to digital ID.

The Digital Decade governance framework is based on an annual cooperation cycle to achieve the common objectives and targets involving both the Commission and Member States. A

[4] https://caiml.org/dighum/dighum-manifesto/

[5] Shoshana Zuboff (2019). The age of surveillance capitalism: the fight for a human future at the new frontier of power: New York: Public Affairs, ISBN 978-1-61039-569.

[6] E.g. must governments impose robot taxation, as they do for workers?

[7] E.g. the use of sensors and cameras to increase the performances of workers.

8 The digital intensity index is a composite indicator derived from the survey on ICT usage and e-Commerce in enterprises.(source EUROSTAT)

review stage of the targets is planned in 2026 to take stock of technological, economic and societal developments.

On WSIS High Level Event 2024 different panels and workshops were held exploring the impact of digital transition on our lives and behaviours. The panel "New Normal: are we ready for it?[9]" considered the global impact of digital transition and related side effects on society including creativity, art and culture. The panel "Al and Global Challenges: Ethical Development and Responsible Deployment[10]" outlined the potential drawbacks due to a massive use of bots mainly trained by western content suggesting the idea to localise AI services in different cultural communities to offer a tailored outcome and preserve differences.

Creativity: human v/s machines?

The digital infrastructure standards foreseen by the Digital Decade outline the relevance of clouds and edge computing. In the process of digital transformation, the "AI & ML" sector seems to strongly impact large part of society involving privacy, freedom, labour, security, lifestyle, and more. Never-before so many public authorities have done so much to introduce regulatory frameworks for a technology.

Nowadays among "platforms" we must include AI application "platforms". The most appreciated by citizens are the generative AI applications that provide on the fly generated content as text, images, and more. In few years these creative "services" gained huge popularity and become an everyday tool to perform our tasks, write a document, better a picture, finalise a video clip. Generative AI (genAI) presents societies with tremendous benefits and promises of progress, but it also has the potential to cause harm in unexpected ways.

Artificial intelligence, expert systems and fuzzy logic were some of the keywords in the 1980s, at that time 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.
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. 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"[11].
The concern was: "artificial intelligence design will improve itself to think faster and deeper, then the improved version would improve itself, and so on, exponentially."
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.
Why were citizens particularly concerned about AI? One of the possible reasons is due to the term "artificial intelligence" that ignited the understanding that there were two "intelligences"

[9] https://www.itu.int/net4/wsis/forum/2024/Agenda/Session/238

[10] https://www.itu.int/net4/wsis/forum/2024/Agenda/Session/199

11 Stanley and Stella: https://youtu.be/3bTqWsVqyzE, last accessed January 2024.

one human and anther artificial potentially competing to rule the world. Some citizens consider this technology as a kind of "pandora box" once opened will not be under human control, self-improve its "intelligence" eventually conflicting with humans. This was many times proposed by Sci-Fi movies as War Games or Eagle Eye[12], just not to mention "2001 a space odyssey[13]". AI systems can provide incredible results analysing on the fly billions of data identifying hidden patterns. This can lead to incredible added value outcome in biology, health, text analysis, and more. The main concern expressed by key players in AI field it is not overcoming human intelligence, they are concerned about the potential misuse of such outputs or potential full dependence from AI outputs. AI systems are not "intelligent" as usually assumed by humans, they can provide a valuable output, but they do not "have the intrinsic knowledge or deep understanding" of that output. Thanks to neural networks, even not supervised, they can provide us the best solution, but they do not have the awareness, deep knowledge, of that solution.

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 [4 – 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 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. 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.

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 systems are not set actually; another 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? Apart from the typical focus on moral and material responsibilities this example depicts even the potentially different "logics" due to different cultural models (eastern / western) outlining the close link between AI outputs and specific training.
More at practical level GPT can support businessman creating an almost perfect contract but this will be generated probably accordingly with the US regulations.

12 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)

13 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.

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 2024 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[14]. This aspect it is not far from the potential bias due to the "mainstream content" delivered online throughout the world via the Internet.

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? 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? Lawyers are already animating the debate together with other stakeholders. "Local content" will be soon generated by "local" bots?
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.

AI for Good v/s AI for Bad

Some years ago, to promote awareness and stakeholders' cooperation the International Telecommunication Union created the already mentioned event named "AI for Good". On May 2024 in Geneve the "AI for Good general conference and exhibit" 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 is equally shared between citizens and AI developers.
If AI for Good can perform different added value activities as an example: Autonomous Systems, Cyber Security, Criminal Investigations identifying Suspects and Criminals using Predictive Analytic Tools, Crime Prediction like the fiction movie Minority Report, Counterterrorism, Cyberwar. One of the recurrent topics on AI for Good conference was about AI for Good V/s AI for Bad – Malicious uses of AI [5 – Ronchi 2021], the fight between AI generated fakes and AI debunked ones? [6 - 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".
AI can be used maliciously affecting cybersecurity, physical security, political security, and more. Malicious use of AI in cybersecurity can include AI powered cyber-attacks, human-like Denial of Service, AI powered malware, detect attack targets based on behavioural data. Automated social Engineering attacks via impersonation of thrusted users, deep fake of voice

[14] Cultural models were already evident in the 1980s - e.g. MS Encyclopaedia.

or video to persuade disclose critical information, automated phishing. Automated vulnerability discovery will be faster and harder to be discovered.
Thanks to malicious use of AI, soon your face can become a trigger for the execution of malware, or even ransomware activated when target found.
This to do not consider the applications in the field of hybrid-and cyber-warfare. There is a huge set of cyberwar related use of AI including lethal autonomous weapons.

AI Governance and ethics

The depicted impact of digital transition or better digital transformation plus the potential specific impact of recent AI developments, their deployments and relative rapid dissemination imposes to focus on AI governance.
AI systems rely on increasingly large amounts of training data, often with personal data collected through various methods, raising policy questions about privacy and ownership. To ensure integrity, we must bring together actors from the AI, privacy, and data protection communities to explore policy responses.

Few years ago, 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.
Today Artificial Intelligence is growing at a phenomenal speed. Mind boggling events like introduction of CHAT GPT-4 in 2024 represent unprecedented challenges. Hence, regulation of Artificial Intelligence including Generative AI becomes of crucial and critical importance for all stakeholders in AI ecosystem.

On the regulatory side overall aim is to cover all AI, including traditional symbolic AI, Machine learning, as well as hybrid systems. Different international organisations and governments are working on regulations. More in detail UNESCO launched the AI Initiative[15] and published different reports on this topic including UNESCO Generative AI, UNESCO Generative Ai in Education[16], and more[17].
IEEE published Ethically Aligned Design[18], 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[19]” or “AI for Safer Children”.
The OECD in 2019 published one of the first document providing some basic guidelines to regulate AI technology “Recommendation of the Council on Artificial Intelligence”[20] that become a key reference globally, including the definition of an AI system provided “Artificial Intelligence (AI) is a general-purpose technology that has the potential to: improve the welfare and well-being of people, contribute to positive sustainable global economic activity, increase innovation and productivity, and help respond to key global challenges. It is deployed in many sectors ranging from production, education, finance and transport to healthcare and security.”

[15] https://www.unesco.org/en/artificial-intelligence/recommendation-ethics?hub=32618
[16] https://www.unesco.org/en/articles/generative-artificial-intelligence-education-think-piece-stefania-giannini
[17] https://www.unesco.org/en/artificial-intelligence
[18] https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9398613
[19] https://unicri.it/topics/Toolkit-Responsible-AI-for-Law-Enforcement-INTERPOL-UNICRI
[20] https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

At the end of 2023, the OECD has decided to play its part, leveraging its unique cooperation infrastructure and convening power to support and foster a positive and proactive message about AI and privacy. The OECD active cooperation across borders, sectors, and areas of expertise is evident. The key role of OECD is to act as an observer and an historical partner of Global Privacy Assembly (GPA) that is a reference global network of Privacy Enforcement Authorities (PEAs) together with the Council of Europe (CoE).
In early 2024, the OECD formally launched the OECD.AI Expert Group on AI, Data, and Privacy, bringing together leading AI and privacy experts worldwide (data protection authorities, policymakers, industry, civil society, and academia). The OECD policy observatory published the 2024 OECD AI principles update[21]. The OECD Ministerial Council Meeting held the same year offered the opportunity to update the principles on AI governance established for the first time in 2019, the already mentioned first intergovernmental standard. The 2024 update considers new technological and policy developments, ensuring they remain robust and fit for purpose. The updated principles now address emerging challenges with an enhanced focus on safety, privacy, intellectual property rights and information integrity.
These principles defined the basis for innovative, trustworthy AI respectfully considering human rights and EU democratic values. Up to now we count 47 countries that adhere to these Principles.
The role of these Principles is to provide recommendations and guidelines to policy makers. This process allows the creation AI risk frameworks, building a foundation for global interoperability between jurisdictions.
The principles are based on shared values and can be summarized: Inclusive growth, sustainable development and well-being (Principle 1.1), Human-centred values and fairness (Principle 1.2), Transparency and explainability (Principle 1.3), Robustness, security and safety (Principle 1.4), Accountability (Principle 1.5).
Coping with these principles we find recommendations for policy makers such as: Investing in AI research and development (Principle 2.1), Fostering a digital ecosystem for AI (Principle 2.2), Shaping an enabling policy environment for AI (Principle 2.3), Building human capacity and preparing for labour market transformation (Principle 2.4), International co-operation for trustworthy AI (Principle 2.5).
As stated by OECD “The Principles serve as a benchmark for responsible AI development and a critical checklist to address these rapid changes effectively, ensuring that AI continues to benefit society without compromising standards and safety.”
In addition, as already mentioned, there are some core aspects such as interoperability of AI and the definition of AI systems and its lifecycle.

What about European Commission initiatives in this field? The EC join Research Centre explored the opportunity AI-generated synthetic data presents for privacy-safe data use.
In such specific field EU Regulations adopt a risk-based approach to regulation. AI should be as neutral as possible to cover techniques that are not yet known/developed.
On 17 May 2024, "the Council of Europe adopted the first-ever international legally binding treaty[22] aimed at ensuring the respect of human rights, the rule of law and democracy legal standards in the use of artificial intelligence systems. The treaty, which is also open to non-European countries, sets out a legal framework that covers the entire lifecycle of AI systems and addresses the risks they may pose, while promoting responsible innovation. This document was the outcome of two years of work due to an intergovernmental body, the

[21] Care of Juraj Čorba, Audrey Plonk, Karine Perset, Yoichi Iida

[22] https://search.coe.int/cm#{%22CoEIdentifier%22: [%220900001680afb11f%22],%22sort%22:[%22CoEValidationDate%20Descending%22]}

Committee on Artificial Intelligence (CAI) composed by 46 CoE members states, 11 non-member states[23], the European union plus the private sector, civil society and academia, who participated as observers. The convention adopts a risk-based approach to the design, development, use, and decommissioning of AI systems, which requires carefully considering any potential negative consequences of using AI systems."
Few days later, on 21 May, the Council of the European Union adopted the EU AI Act, which once published in the EU Official Journal in June, became the first set of AI regulations that has undergone a full legislative approval process.
The EU AI Act is structured in 113 articles and counts 13 annexes, its holistic risk-based approach is suitable for any player in the field of AI from developers to deployers.
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.[24]"
Dealing with a fast evolving technology the key aspect in defining a framework is to find a golden balance between shaping the evolution and leaving to technology the freedom to evolve naturally.

Recently the Australian Government has realized version 1.1 of its "Policy for the responsible use of AI in government[25]"
If we focus on Asia to approach the Asian perspective, there is an interesting contribution by Denise Wong[26] together with co-authors in "A fast-evolving AI and privacy policy landscape calls for international cooperation" concerning the need to cooperate internationally in data governance and privacy to face the growing consensus and fast-evolving AI and privacy policy landscape. "In this complex and fast-paced environment, where countries and organisations compete to reap the full benefits of AI, legitimately protecting the rights of some, including their privacy and intellectual property rights, can be seen as affecting data quality and availability for training AI models, thus frustrating innovation capabilities." The paper remarks that privacy regulations could hinder AI and data-driven innovation. To mitigate this problem privacy community, regulators and professionals are actively cooperating to implement privacy-friendly AI and even use AI in the service of privacy. Data Protection Authorities (DPA) or Privacy Enforcement Authorities (PEA) decided to create AI departments to strengthen their expertise in AI systems, improve the understanding of privacy risks and lastly implement the EU AI Act.
On May 2023 CNIL (French DPA) published a document entitled "Artificial intelligence: the action plan of the CNIL - In the face of recent news on artificial intelligence, and in particular so-called generative AIs such as ChatGPT. The CNIL published an action plan for the deployment of AI systems that respect the privacy of individuals.[27]"
On 15 March 2023 the UK ICO issued and update of a comprehensive guidance on AI and data protection, later they launched a consultation series on generative AI and data protection[28].
Singapore is launching a set of guidelines on how personal data can be used to train and develop AI recommendation and decision-making systems.

[23] Argentina, Australia, Canada, Costa Rica, the Holy See, Israel, Japan, Mexico, Peru, the United States of America, and Uruguay

[24] Artificial Intelligence for Europe

[25] https://www.dta.gov.au/blogs/responsible-choices-new-policy-using-ai-australian-government#:~:text=To%20protect%20Australians%20from%20harm,of%20the%20policy%20effect%20date.

[26] Denise Wong - Assistant Chief Executive, Data Innovation & Protection Group - Singapore Infocomm Media Development Authority (IMDA) https://oecd.ai/en/community/denise-wong

[27] https://www.cnil.fr/en/artificial-intelligence-action-plan-cnil

[28] https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ico-consultation-series-on-generative-ai-and-data-protection/

Since 2023 the number of technological and organisational initiative to face these issues is increasing. Among them, France, Norway and Singapore (started already in 2020) creating an increasing number of sandboxes applied to AI and fostering the involvement of several data intermediaries to promote a responsible data sharing within AI economy.
One of the key aspects and more debated topic is privacy preservation in the AI domain. There are interesting applications in this field developed by research teams from École Polytechnique Fédérale de Lausanne (EPFL)[29], the University of Michigan[30] and the University of Wisconsin[31]. These research centres developed the AI ability to read and interpret privacy policies of online services thanks to deep-learning program.
An additional opportunity to push forward the research and achievements is to foster he cooperation between AI and the privacy communities on the initiative Privacy Enhancing Technologies (PETs)[32].
Moving the focus on South Africa, the South African National AI Framework[33] recently affirmed its commitment to ethical AI development and use. This it is not a hard law, but a set of guidelines to ensure AI systems are transparent[34], accountable and designed to promote fairness while mitigating biases. These guidelines include[35]:

- the establishment of robust data governance frameworks to protect privacy and enhance data security, and
- the development of standards for AI transparency and explainability to foster trust among users and stakeholders.

« The Department of Communications and Digital Technologies acknowledges the need for developing the National AI Policy for the Country. The department has done extensive research over the past couple of months and has developed an AI Policy Framework. This Framework unpacks the pillars we want to focus on during this policy development. We therefore request comments on this document as we will be doing formal consultations during the month of August and September. »

AI regulation and governance is not only promulgated at international and national level, in many cases it also involves policy being developed and implemented at local level. Following this trend, we find recent developments in the U.S. (e.g. New York and California). “In the US, action has predominantly occurred at the state-level, because the federal government has yet to move from general frameworks and recommendations to enacting federal legislation about artificial intelligence.[36]”. New York surpassed California in proposing AI legislation! And then, again, California has been more successful at enacting legislation[37] (as did Oregon, Utah, Maryland, Illinois, and Florida).

[29] https://actu.epfl.ch/news/opening-of-the-new-epfl-ai-center-a-hub-for-ai-i-4/

[30] https://ai.engin.umich.edu

[31] https://rise.wisc.edu/rise-ai/

[32] https://ico.org.uk/media/about-the-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf

[33]https://htxt.co.za/wp-content/uploads/2024/08/South-Africa-National-AI-Policy-Framework.pdf

[34]https://www.dcdt.gov.za/images/phocadownload/AI_Government_Summit/National_AI_Government_Summit_Discussion_Document.pdf

[35] https://www.dcdt.gov.za/sa-national-ai-policy-framework/file/338-sa-national-ai-policy-framework.html

[36] Hemant Bhargava, Ayan Mani, Anirudh Murugesan – “Artificial Intelligence Regulation in the United States” - July 19, 2024 - Center for Analytics and Technology in Society (CATS), UC Davis

[37] https://www.linkedin.com/pulse/week-ai-california-weighs-regulations-openais-reset-sara-bloomberg-81qmc/?trackingId=2AZQpfbWRLuDQVS2Ce8ijA%3D%3D

Final remarks

In conclusion, besides the basic question of « which states (or any local government entity) are in pole position in AI regulation », several essential questions deserve to be posed:

Is there a 'copycat' effect in the legislation that is produced, i.e. are the less active states piggybacking on the work of other more active states in forming their own AI regulation? If so, what are the pro and con implications of such copycat behaviour?

What factors are driving the dominance of « government's own use of AI » in AI-related policymaking? (This raises the question of the willingness or capability of governments to regulate beyond their own corner, i.e. to include the private sector.)

What are some trends that exist with respect to state-based regulation? Are there, within the states, any political correlations with characteristics of industry ecosystem and private enterprise? (e.g., do states with heavy concentration of finance and banking follow a different approach vs state that are dominated by healthcare or by manufacturing?)

How are or should be articulated state/local-level AI policies with national AI policies (not to speak here of regional/international policies!)? Do synergies exist and say to create? Are disparities and conflicts possible, and how to mitigate them?

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 collection and reuse of information. 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 the design or co-creation of cyber technologies always keeping humans in the focus.

References

[1.] Schwab Klaus. "Shaping the Fourth Industrial Revolution", World Economic Forum 2016

[2.] V International Conference "Tangible and Intangible Impact of Information and Communication in the Digital Age" to be held in Khanty- Mansiysk, Russian Federation, on 6-8 June 2023 within the XIV International IT Forum with BRICS and SCO participation and the UNESCO Information for All Intergovernmental Programme (IFAP)

[3.] Ronchi Alfredo M., (2020). Digital transformation, proceedings ICCC New Delhi, Cyberlaw ISBN:978-0-385-50386-0

[4.] Stuckelberger Christoph, Duggal Pavan (2018), Cyber Ethics 4.0: Serving Humanity with Values, ISBN 978-88931-265-8, Globethics.net

[5.] 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

[6.] European Union (2016) Joint Framework on countering hybrid threats a European Union response, 2016