Strengthening democratic engagement through value-based generative adversarial networks

D4.1 SOLARIS
Interdisciplinary mapping of political risks and implications of GANs infodemic

Lead Author:
Calogero Caltagirone, Angelo Tumminelli (Lumsa University)
With contributions from:
UvA, UM, DEX, UC3M, ANSA, AIIS, BMB, AMI, CSA
Reviewers: UNEXE, ECSA

Deliverable nature:	R-Document
Dissemination level: (Confidentiality)	PU
Delivery date:	May 2024
Version:	Final
Total number of pages:	30
Keywords:	Geopolitical risks, Infodemics, Deepfake diffusion, Comparative Law frameworks, GANs impacts on democracy.

Executive summary

This *Deliverable* investigates the political risks and negative implications due to the circulation of Generated Adversarial Networks (GANs) technology by presenting, in particular, an interdisciplinary mapping of the geopolitical consequences, both at a national and international level, connected to the infodemic. This mapping process serves to define some strategic elements in understanding the infodemic phenomenon: first of all, the *Deliverable* identifies the political actors involved in the diffusion and circulation of GANs, highlighting their interests for the purposes of controlling public opinion and the ideological orientation of the users. Furthermore, the impact of the infodemic in individual state communities and its effects on the international geopolitical structure is presented.
In fact, the spread of deepfakes is not only able to influence the exercise of democracy in individual states by conveying ideological thought and orienting the political consciences of citizens but, indeed, it can influence international dynamics by producing conflicts and fueling the polarization of points from a political point of view. Therefore, the *Deliverable* presents not only the risk analysis but also wants to highlight the need for responsible use of GANs technologies to exercise democracy freely and free from ideological conditioning. Through the use of an interdisciplinary methodology and transdisciplinary approaches, the *Deliverable* also intends to show a framework of comparative law in order to understand the strategies of each individual state to combat infodemic risks and allow its citizens to freely orient their political consciousness.
With the contribution of experts from the journalistic sector and international diplomacy, we will try to understand what type of networks can be interrupted - but also created - by GANs and who could benefit from synthetic virtual agents that share political content.
To introduce the work, it should first of all be underlined that the ability of AI to rapidly analyse a significant amount of data, to recognize patterns and to create predictive models starting from existing knowledge, makes it a powerful tool for the development of humanity, especially for the sustainable one. For example, AI can be used to improve the responsible conduct of businesses, facilitating the promotion of human rights in individual political communities. However, the misuse of AI can negatively impact several fundamental rights, reshaping social dynamics and impacting global democratic systems. It can therefore transform the current geopolitical reality by altering the life of democracy. According to the Council of Europe, two key principles govern democracy: individual autonomy, which consists in the idea that no one should be subject to rules imposed by others; and equality, which means that everyone should have the same opportunities to influence decisions that affect people in society (Council of Europe, 2023). These principles are also human rights values. Article 21 of the *Universal Declaration of Human Rights* tells us that we have the right to participate in government. However, being able to enjoy this right means enforcing other rights, such as freedom of thought, conscience and religion, freedom of expression, freedom of peaceful assembly and association and the right to privacy. The connection between human rights and democracy is fundamental. Therefore, the influence of AI on democracy is directly proportional to the protection/violation of some human rights.
Freedom of thought is one of the main rights of a democracy: people must be able to think freely without being punished for it. This condition creates the pluralism which is a pillar of a democratic society. Artificial intelligence systems have the power to stimulate human creative thoughts, presenting concepts that some may not have considered. However, they are also capable of showing only the content a person wants by recording their previous online behavior, encouraging confirmation bias instead of facilitating their critical thinking. Thinking critically about our surroundings is essential to having pluralistic visions and inclusive debates. AI can even create fake, realistic videos, audio, and images that can challenge decision-making and be used as propaganda to influence public opinion and manipulate elections.
In a democracy, it is not only essential to be able to think critically and formulate your thoughts from pluralistic and reliable sources, but it is also essential to express this opinion and be able to come together to discuss with others. Furthermore, to do this, it is also important that individuals feel that they have an area of autonomous development, interaction and freedom, a "private sphere" with or without interaction

with others, free from state intervention and excessive non-intervention. requested of other uninvited individuals and to determine who holds information about them and how it is used.
Therefore, democracy should not undermine the right to privacy, which affects the freedom to create one's own thoughts, to express one's opinion with others without someone "spying" on it.
The use of AI for surveillance can significantly improve public safety strategies, to the benefit of the community. However, authorizing the legal use of AI for mass surveillance, without specific limits, hinders the right to privacy and shapes the way people feel free to behave, to speak, to present their positions through interest groups or to come together to protest against decisions they do not agree with.
If you carry out a comparative study of the various national jurisdictions, you understand that AI laws propose a risk-based approach, in which legislative intervention is customized based on the level of risk. It is possible to distinguish between AI systems that pose unacceptable risks and will be banned; those at high risk that will be regulated; limited risk models that will only require transparency commitments; and those with low or minimal risk which will involve voluntary codes of conduct. AI applications would be regulated as strictly necessary to address specific levels of risk. This approach seems well suited to enabling differentiation and avoiding over-regulation that could interfere with the innovation and competitive advantage of European AI companies compared to their peers outside the borders of the European Union . Indeed, the risk-based measure is generally welcomed by stakeholders, the Council of the EU and the European Parliament. In practice, there will be common mandatory requirements applicable to the design and development of AI systems before they are placed on the market and the way in which ex-post controls are carried out will be harmonised. However, it is important to draw attention to the ex-ante risk assessment. This is left to the programmer who will have to explain the specific design ethics of the system, such as the logic of the AI system and algorithms, and the assumptions made about the recipients of the system.
In conclusion it can be said that artificial intelligence systems can undoubtedly be useful for human development. Every day human beings make important decisions: if they are managers, they decide who to hire; if they are CEOs, they decide how to conduct business in unregulated or rule-free environments; if they are statesmen, they decide how to manage a country based on their political values and their electorate, and so on. Managers can decide to hire someone who is like them or socially recognized as "suitable"; CEOs may decide to conduct irresponsible practices to maximize their profit; Politicians may seek re-election by making short-term decisions and not worrying about long-term consequences.
AI can be a gentle nudge to direct humans towards responsible choices, if programmed to structure a good choice architecture that allows governments to protect citizens' freedom by encouraging them to make wiser decisions. However, AI can also be a dangerous stimulus influenced by biased and unregulated algorithms. For this reason, multilateral cooperation is key to creating an environment of deterrence and responsible AI.
Overall, within its limits, the current EU regulatory framework would facilitate the benefits of AI by enabling trust, seeking to minimize harm to fundamental rights and democracy through a refineable approach to risk management and leveraging the potential to improve human development. However, we must ask ourselves what further regulatory indications are necessary for the regulation of AI and to avoid the negative risks of the infodemic.
In this sense, the interdisciplinary work of mapping the geo-political risks connected to the infodemic and the use of GANs technologies presented in this *Deliverable* meets the need to define ethical, regulatory and political strategies to stem the negative consequences and promote, instead, the exercise of a free and responsible political democracy in which the contribution of individual citizens contributes to the common good and the realization of universal peace.

Figure 1. The topic of Deliverable

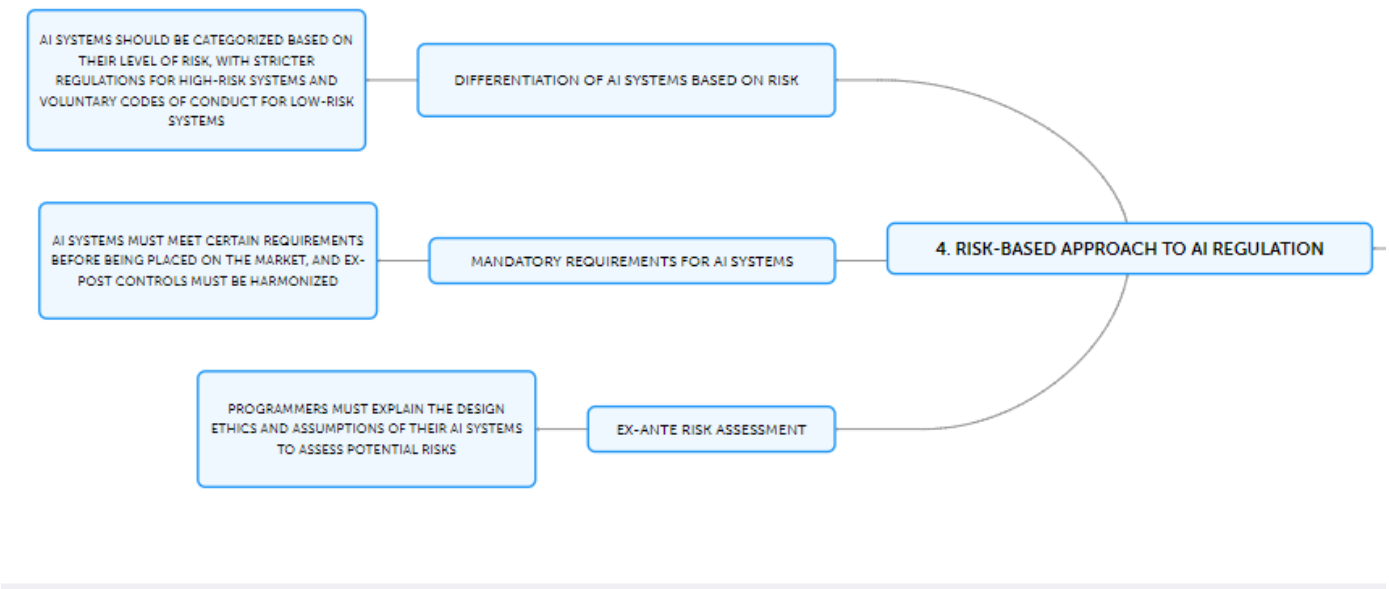

Figure 2. Risk-based approach to AI Regulation

Document information

Grant agreement No.	**101094665**	**Acronym**	**SOLARIS**
Full title	**Strengthening democratic engagement through value-based generative adversarial networks**		
Call	**HORIZON-CL2-2022-DEMOCRACY-01**		
Project URL	**https://projects.illc.uva.nl/solaris/**		
EU project officer	**Ms. I. von Bethlenfalvy**		

Deliverable	**Number**	4.1	**Title**	Interdisciplinary mapping of political risks and implications of GANs infodemic
Work package	**Number**	4	**Title**	Designing regulatory innovations for infodemic risks mitigation
Task	**Number**	T4.1	**Title**	Interdisciplinary mapping of political risks and implications of GANs infodemic

Date of delivery	**Contractual**	M16	**Actual**	M16
Status	version 1		☒Final version	
Nature	☒Report ☐Demonstrator ☐Other ☐ORDP (Open Research Data Pilot)			
Dissemination level	☒Public ☐Confidential			

Authors (partners)	LUMSA, UvA, UM, UNEXE, DEX, UC3M, ANSA, AIIS, BMB, AMI, CSA			
Responsible author	**Name**	Angelo Tumminelli, Calogero Caltagirone		
	Partner	Lumsa University	**E-mail**	a.tumminelli@lumsa.it c.caltagirone@lumsa.it

Summary (for dissemination)	This Deliverable investigates the political risks and negative implications due to the circulation of GANs technology by presenting, in particular, an interdisciplinary mapping of the geopolitical consequences, both at a national and international level, connected to the infodemic . This mapping process serves to define some strategic elements in understanding the infodemic phenomenon : first of all, the Deliverable identifies the political actors involved in the diffusion and circulation of GANs , highlighting their interests for the purposes of controlling public opinion and the ideological orientation of the users. Furthermore, the impact of the infodemic in individual state communities and its effects on the international geopolitical structure is presented.
Keywords	Geopolitical risks, Infodemics, Deepfake diffusion, Comparative Law frameworks, GANs impacts on democracy.

Acknowledgement

Funded by the European Union

Disclaimer

Table of contents

List of figures

Abbreviations and acronyms

Abbreviation	Meaning
AI	Artificial Intelligence
CA	Consortium Agreement
CC BY 4.0	Creative Commons Attribution 4.0 International
Celeb-DF, DFD, DFDC	Publicly available datasets developed for deepfake detection
CERN	European Organisation for Nuclear Research
DDI	Data Documentation Initiative
DMP	Data Management Plan
DOI	Digital Object Identifier
DPO	Data Protection Officer
EC	European Commission
EU	European Union
FAIR data	Data which meet principles of findability, accessibility, interoperability, and reusability
GANs	Generative Adversarial Networks
GDPR	General Data Protection Regulation
UC	Use Case
WP	Work Package

1 The political actors involved in GANs design and diffusion, and their interests

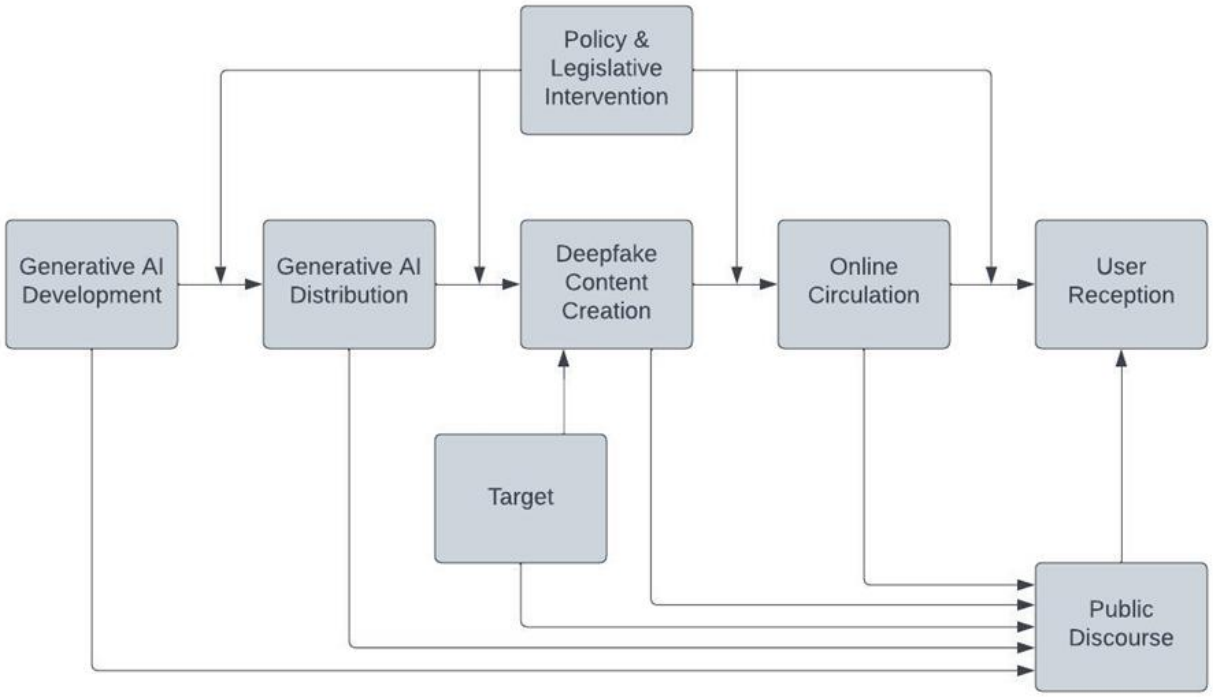

Figure 3. Approximation of the deepfake actor-network

Building on the deepfake actor network presented in SOLARIS Deliverable 2.2 (approximation shown in Figure 1 and expanded view in Figure 2) this section identifies the explicit political dimensions of the significant social actors involved in production, distribution and reception of AI-generated content. This involves understanding how political values are established or introduced by different social actors and mapping how these values may then spread throughout the rest of the network impacting the actions of other actors. This mapping will help identify points of intervention wherein policy and legislation may be introduced to counteract the spread of anti-democratic ideas and/or promote pro-democratic ideas through AI-generated content. The following section is separated into sub-sections focusing on the different groupings of social actors established previously in D2.2 and identifies how major political values are introduced during each phase of the pipeline with reference to relevant literature.

1.1 Generative AI development

Bias, inaccuracy and censorship: the creation of training datasets for generative AI currently involves human programmers making certain explicit and implicit decisions about what content to include or exclude and how that content is arranged. These decisions involve inherently political decisions and thus instil political values in generative AI programs. For example: if the dataset for “criminal” is intentionally or unintentionally populated with images of Black men, when prompted to produce images of criminals the generative AI program will likely return images of Black men and thus reinforce racial biases. Similarly, the censoring or exclusion of certain data in these datasets will limit the representation of certain groups of people in generative AI outputs. Bias, inaccuracy and censorship ultimately undermine democracy by promoting distrust and division between communities of citizens and by contributing political disengagement from underrepresented communities. For example, bias can be seen as a systematic form of inaccuracy.

- **Value-sensitive design for AI:** outside of legislation and policy, numerous national and international partnerships and initiatives have emerged in recent years advocating for human-centric AI development and outlining various voluntary guidelines, frameworks and principles that are necessary to uphold human rights and dignity. Examples of these initiatives include the OECD Principles on Artificial Intelligence, UNESCO's Recommendations on AI Ethics, and NIST's AI Risk Management Framework. Through value-sensitive design, there have been efforts to implement these initiatives as practical and material design processes and business practices. Through Variable Speed Drive (VSD) techniques such as bias mitigation, developers are embedding political values into generative AI programs such that the content they generate will be reflect these values and thus influence other social actors in the network.

- **Readings:** *Atlas of AI* (Crawford, 2021), *Discriminating Data* (Chun, 2024), *Artificial Whiteness* (Katz, 2020), "On the Dangers of Stochastic Parrots" (Bender *et al.*, 2021), "Designing AI with AI4SG Values" (Umbrello and Van De Poel, 2021), "Designing AI With Human Rights Values" (Aizenberg and Van Den Hoven, 2020).

1.2 Generative AI distribution

- **Advertising:** marketing material for generative AI products encourage or discourage particular uses of this technology that may entail political values. For example, advertising a generative AI app as a tool for producing non-consensual deepfake pornography of women will perpetuate misogynist ideas.

- **Hype, exaggeration and deception:** many tech companies have been accused of misrepresenting AI products by exaggerating their technical capabilities or by falsely asserting that these technologies are entirely objective in their outputs. Such advertising practices encourages users of AI technologies to do so uncritically which could lead to inappropriate deployment and the perpetuation of those biases and inaccuracies introduced during development.

  Diffusion models: also known as denoising diffusion probabilistic models (DDPMs), diffusion models are generative models that determine vectors in latent space through a two-step process during training. The two steps are forward diffusion and reverse diffusion. The forward diffusion process slowly adds random noise to training data, while the reverse process reverses the noise to reconstruct the data samples. Novel data can be generated by running the reverse denoising process starting from entirely random noise.

- **Readings:** *Deceitful Media* (Natale, 2021), "Talking AI into Being" (Bareis and Katzenbach, 2022).

1.3 Content creation

- **Politically motivated disinformation:** generative AI programs can be used as tools for the production of convincing deepfakes of political figures and public officials so to promote a particular narrative about political issues or to undermine figures and/or the institutions they represent. Of particular note are far-right extremists and bad actors backed by foreign countries.

- **Ideological representation:** beyond specific political issues, content creators may intentionally use generative AI to create content that promotes a particular ideology or worldview. Notably, this includes imagery that misrepresents certain groups or communities (e.g., deepfake pornography presenting women as sexual objects).

- **Visualisation:** generative AI programs may be used to produce artificial images that better illustrate particular political points. This includes AI-generated images that shed new light on historical events (e.g., *Dimensions in Testimony* exhibition features deepfakes of deceased victims of the Holocaust telling their experiences; *Exhibit A-i* exhibition features AI-generated images of previously

unrecorded events at the Manus Island and Nauru immigration centres). This may also apply to speculative imagery to emphasize a particular political point (e.g., Images from the *ThisClimateDoesNotExist* project showing landmarks in extreme environments to visualize the effects of climate change; US Republican party advertisements showing crowds of migrants crossing the US-Mexico border).

- **Readings:** "Truth, Lies and Automation" (Buchanan *et al.*, 2021).

1.4 Targets

- **Target's political position/power:** political figures and government officials that are often targeted by deepfake disinformation are often associated with particular ideologies, movements and political parties/organizations and so deepfakes of these figures may intentionally or unintentionally draw upon these associations. For example, a deepfake of a political figure may not only misrepresent this individual's views or actions but further misrepresent their organization or associated ideology to viewers as a result.

1.5 Online circulation

- **Censorship, polarization and neighbourhoods:** the technical structures or online information networks dictate the flow of information and cluster users into neighbourhoods of individuals linked by similarity (e.g., race, sexuality, gender, political opinion). Within these neighbourhoods, particular political values may be promoted while others are discouraged creating political echo chambers that filter out dissimilar or critical content. Within these insular communities, hatred of those dissimilar is masked as love of those similar encouraging polarization between different communities and enabling extreme political views to spread freely and with greater impact via strong interpersonal ties. As a result, politicized AI-generated content circulating within these neighbourhoods may spread more quickly, be received with less critique and may have a greater impact on users.

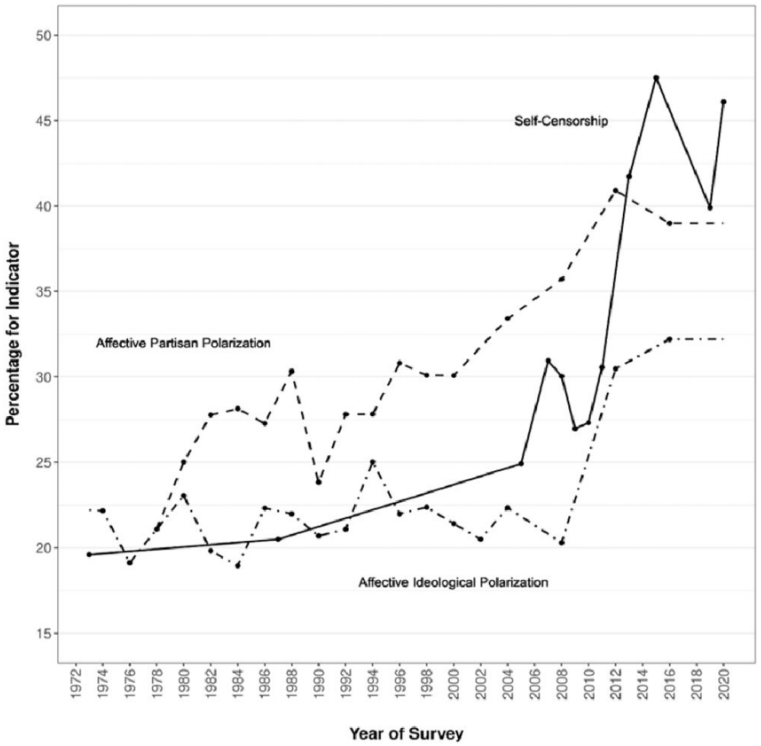

Figure 4. The Relationship Between Self-Censorship and Affective Polarization

- **Censorship and content moderation:** to stem the flow of disinformation online, social media platforms have implemented community standards and policies governing what content can be posted and shared among users, as well as actively removing disinformation content or labelling it as such. Such efforts may be effective in limiting the spread of AI-generated political disinformation and the values such content promotes. However, these policies have been seen as violations of users' freedom of speech and may have the opposite effect while also encouraging users to migrate to other less regulated online areas.

- **Readings:** *Discriminating Data* (Chun, 2024), *Media Manipulation and Disinformation Online* (Marwick and Lewis, 2017).

1.6 Public discourse

- **News agenda:** limiting this discussion to the role of media organizations in spreading AI-generated political content, it is important to acknowledge the role of such organizations in identifying and debunking disinformation, while others, notably smaller organizations with less resources, may intentionally or unintentionally share AI-generated disinformation and thus perpetuate the political narratives such content expresses.

- **AI hype:** misunderstanding and deceptive marketing practices of AI developers have recently led to a so-called "hype" that misrepresents the technical capabilities of such technologies and often frames these products as politically neutral. Such narratives advanced in media reports may encourage uncritical reception of AI-generated media.

- **Readings:** "The AI Doctor Will See You Now" (Bunz and Braghieri, 2022) and *Deceitful Media* (Natale, 2021).

1.7 User reception

- **Misunderstanding and uncritical reception:** as generative AI is a relatively new technology and one that is quickly developing, many users may not appreciate that such technologies can display political bias and so may take an uncritical approach toward AI-**generated content. Furthermore, users may be convinced by AI marketing and media hype so as to believe these ostensibly objective and accurate technologies are knowledgeable and truthful about human affairs. This may mean that users may not recognize AI-generated content and may be more receptive to the political ideas expressed by it, particularly relating complex and nuanced political issues.**

- **Readings:** "The Radicalization Risks of GPT-3" (McGuffie and Newhouse, 2020), *AI Ethics* (Coeckelbergh, 2020), and "Democracy, Epistemic Agency and AI" (Coeckelbergh, 2022).

1.8 Policy and legislative interventions

- **Political spheres of influence:** with the increasing popularity of generative AI, national and international regulations and legislation governing the development of such technologies have begun to emerge exerting external political influence on developers. While individual countries have begun introducing national strategies and policies, major spheres of political influence have emerged from the US, EU and China that will likely dictate policy approaches in other countries. While the market-driven approach taken by the US allows for generative AI developers to self-regulate in order to boost economic growth, the approaches taken by the EU and China may impact development. China's state-driven approach seeks to dictate the political values communicated in AI-generated content, for example by holding generative AI developers accountable for harmful content produced by their products including content that is critical of the state. Meanwhile, the EU's approach is rooted in preserving fundamental human rights through strict regulation and legislation requiring transparency,

human oversight, accuracy and robustness in AI development. Notably, the EU AI Act details specific transparency and disclosure requirements for generative AI programs including explicit labels for AI-generated content and a means by which such content can be easily detected. With strict regulation already in place in Europe and with developers facing harsh penalties for violations, a "Brussels Effect" is expected in which these regulations are adopted globally thus encouraging EU political values (e.g., bias mitigation) to be implemented as standard.

- **Readings:** "The Race to Regulate" (Bradford, 2023), "The Chinese Approach to Artificial Intelligence (Roberts *et al.*, 2021), "Generative AI in EU Law" (Novelli *et al.*, 2024), "Cutting Through the Hype" (Weikmann and Lecheler, 2023), and EU AI Act.

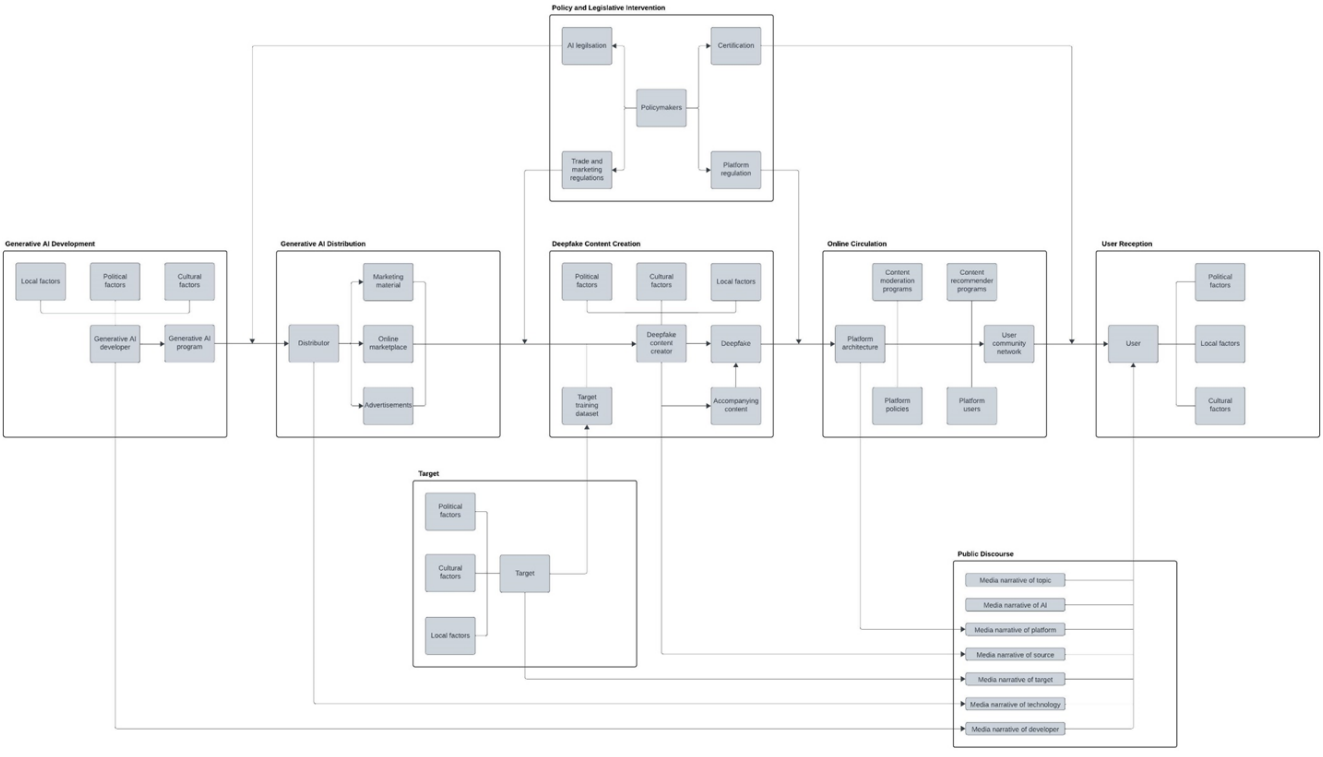

Figure 5. Expanded view of the deepfake actor-network

2 The impacts on political communities

Recent developments in the use of AI in different areas and the breakthrough of ChatGPT and other chatbots based on AI, have raised awareness of the technology's political, economic, financial and geopolitical consequences. Each new development raises concerns on how AI could be used for beneficial means but also how it can be used to harm political processes, international relations, financial systems, etc. When it comes to the impact of AI on politics and political communities, most of the discussions have been on how AI and Generative Adversial Networks (GAN) can be used to produce fake images and videos and how these can affect the outcome of elections or the present and future of politicians. Also, a lot of focus has been on how foreign government, especially those with authoritarian tendencies, can use AI to spread disinformation in other countries, which may harm political processes and create chaos and distraction among voters.

The 2016 Russian disinformation campaign which affected the US Presidential elections that year, is still fresh in the memory of many and since then the technology has advanced considerably. "Astroturfing"—the use of fake online accounts to give the illusion of support for a policy— has been known for a long time now. But these threats are immediate, day to day and the protection mechanisms have grown bolder and better since a few years ago.

Many of the public discussions about AI and its impact on the political processes are unfocused and hover between fear and enthusiasm. AI is a technology that is still unfolding. Hence, there are inherent limitations to predicting the exact form the technology might evolve to assume in advanced stages. Phenomena that are visible today maybe become irrelevant in the near future and other phenomena related to AI may spring up and have a deep impact in the political processes, which can't be foreseen now. And the fact is that as AI develops, people tend to be aware of its implications and possible risks. Therefore, this may limit considerably the risks that AI development may have inside a political community.

2.1 AI and democracy

Probably the biggest concern related to AI and politics, is the potential impact it could have on the democratic processes. AI is often seen as a threat to society, human life and also democracy. This idea is based on a largely imaginary artificial general intelligence (AGI), able to autonomously perceive, reason, decide and act in various contexts of human affairs, including political processes and democracy.[1] This potential ''AI center'' could turn elections around, disrupt the functioning of governmental institutions, create massive disinformation campaigns based on fake videos and texts which will put forward preferred candidates, or it can shape electoral thinking and attitudes through very sophisticated PR campaigns.

In fact, AI as it is used until now, doesn't have any resemblance at all to this artificial general intelligence. Actually, existing AI is predominantly narrow AI trained on domain-specific data to perform domain-specific tasks. (M. Mitchel, 2019, pg.45) In order to understand how democracy is affected by AI, we need to understand how AI is used for the moment in this narrow context, and secondly, to understand what it can happen in the future, we need to examine if there exists a possibility of AGI coming to life in the near future.

One of the pillars of democracy is that governments should be chosen by those they will serve. But in order for a voter to make an informed decision, he or she needs first of all to be informed and information space has been changing considerably in the last years because of the impact of AI in it. These changes have affected political communities. They have become more heterogenous because of the development of social medias and the use of AI in them. They take information from different sources, most of the time, with no relation with each other and furthermore in direct opposite with each other. This means that for every voter in a political community, there is one distinct opinion about the same issue. Especially in countries with a tradition of several political parties vying for power, this tendency may create a more fragmentized political community, where no party or group of parties is able to establish power of government and therefore create a paralyzed political community. Previous studies have shown that there is a link between homogeneity of communities (in terms of income and race) and the level of social capital: more homogenous communities have higher level of social interactions leading to more social capital (2019). While this homogeneity was based on race and income, in our case we can define heterogeneity in terms of information received and political positions. Numerous sources of information, combined with the AI potential to create undistinguishable fake content, may create the conditions for heterogeneity of opinion on political issues to prevail and may hamper the creation of socio-political groups with similar ideas and opinions inside a political community. This way the political community

breaks down into atomized individuals who don't have much relation with each other in terms of political positions. This phenomenon may be more pronounced in Western democracies, where the religious and ethnic bonds inside the population have dissolved, while in countries where the religious and ethnic ties are still very strong and determinant parts of the collective identity, this heterogeneity of information can't dissolve the homogeneity of religion and ethnicity.

Political communities in democracies have also become less informed rather than more. The fact that there is an abundance of information available for everyone to access it, it meant in the beginning that electorates would become more informed, but with the growth of social media and the use of AI to generate videos, photos and texts, the opposite effect has happened. AI, until now, has helped to make information unreliable and untrustworthy. Voters inside a political community are not sure anymore which information is true and which is not. Rather than push for more engagement with political processes, AI could help make people less prone to be engaged in political processes, because there is a lack of trust on the information received and if the voter is unsure about the information he or she receives, then he or she is unsure to participate in a political community. Generative AI, with its power to generate deepfake images and video, which can put people into situations they've never been in, manipulate the perception of events, and potentially gaslight an entire section of the electorate into believing things which never happened (Fyler, 2023). What can have an impact is not the fact that AI generated content creates fake information and tries to influence voters through it, but the fact that for a voter the idea that AI generated fake content is now common, tend to create mistrust towards all content and therefore a less informed voter and less willing to participate in a political community. Surveys show that almost half of respondents could not tell the difference between real and manipulated videos, with the proportion significantly higher among the older generation (Helmus, 2022), which in many countries constitutes the most actively engaged part in political processes.

At the same time, political communities have become more polarized between usually two opposite positions. In this case, AI is used to generate fake content in order to reinforce the arguments of which side. People usually seek out that content which confirms their previous thoughts and ideas, but in the pre-social media and pre-internet world the amount of content and its quality to reinforce previous ideas, was relatively confined and slow to be distributed. The social media development created an explosion of this kind of content and made its distribution almost instantiations. The utilization of AI to create this kind of content, has improved considerably the quality of it and the power to convince others who see, hear or read it. This improved quality of fakeness or distorted information created by AI, can help increase the political polarization inside a political community. People become more convinced in their beliefs because of the plausibility of videos, text and photos created by AI and are less prone to change their previous ideas and less prone to change political sides inside a political community.

The dangers coming from the use of AI, described above, can't be totally ascribed to AI, but mostly to human nature and how it sees political engagement, democracy, elections, receiving of information, etc.

2.2 AI and authoritarian systems

The much fantasized and feared artificial general intelligence (AGI), an AI center able to coordinate autonomously of human intelligence and able to direct and control humans at some point, is more possible to be implemented in an authoritarian system than a democratic one.

First of all, we have to define what we mean when we say ''authoritarian system''. While there is a wide literature on this topic, the main features of an authoritarian system are defined as follow:

- Free and competitive direct elections to the legislative power or executive power, or both
- Minimal political mobilization and suppression of anti-regime activities

Authoritarian regimes, especially China, have been keen to use AI to reinforce the authoritarian nature of their regimes. For example, Chinese firms have built software that uses artificial intelligence to sort data collected on residents, amid high demand from authorities seeking to upgrade their surveillance tools (Baptista, 2022). China has been using similar technologies for surveillance purposes for a long time, but AI is helping China

to create more detailed profiles of its citizens, even in aspects which were difficult or imprecise with the current technologies.

As in the above case of democracy, even in the case of authoritarian regimes, the discussions about the potential use and impact of generative AI are speculative. While in many Western democracies there is a tendency to worry about the potential of AI controlling humans at some point, in authoritarian countries like China, there is generally a lack of what can be called AI "doomerism" discussion. Whether at major conferences or in academic research circles or private chat groups, existential risks scarcely feature as a major concern in China's extensive AI community (Xiang, 2023).

Authoritarian regimes tend to have a more pragmatic view of generative AI and its uses. They tend to use it for state purposes, usually to reinforce existing regimes and for better controlling their populations. Gulf countries, for examples, which are authoritarian regimes, have become champions of disinformation lately (Jones, 2022). Their disinformation campaigns, supported by generative AI aim not just to control their populations, but also to disrupt the activities of other Gulf countries. Authoritarian regimes, in this case, have a free hand in these activities, because contrary to democracies, there is no public opinion to pressure them.

Generative AI and new technologies in general, are making authoritarian regimes even more authoritarian. While authoritarian government have always tried to repress their population and suppress any dangerous ideas, generative AI may help them to create a totally manipulated society, especially with a new generation which will be born and raised in AI manipulated society, it will not know an alternative.

As well as cracking down on supposedly threatening information or narratives, authoritarians aggressively impose their own narratives to shape public perceptions – often successfully. Where citizens across the region seek to draw attention to popular grievances or state failure, they can be overwhelmed by state-led communications campaigns, which pump out disinformation and fake news, often adopting a violent and misogynistic tone (Lynch, 2022).

This process helps turn social media platforms into a hostile environment in which it is hard to distinguish between propaganda and genuine comments. The resulting lack of certainty creates distrust in the public sphere and has a chilling effect on users, leading many of them to abandon the platforms altogether. Women are particularly liable to be targeted. This disinformation campaign against particular users of social medias to advance ideas against the government, is particularly powerful and successful when it is directed by the government itself.

Gulf countries, in particular Saudi Arabia and United Arab Emirates are “digital superpowers” partly because of their investments in controlling the online agenda through a variety of means, including troll and bot armies, as well as astroturfing – the practice of creating the false impression that legitimate users are spontaneously supporting a cause (Jones, 2022).

The much feared artificial general intelligence (AGI) is much more probable to be created in an authoritarian regime, but in difference to the fears in democratic countries, it will not be controlled by AI but by humans controlling AI in authoritarian regimes, with the purpose of creating a totally manipulated and controlled society, in which the threats to the regime will be eliminated.

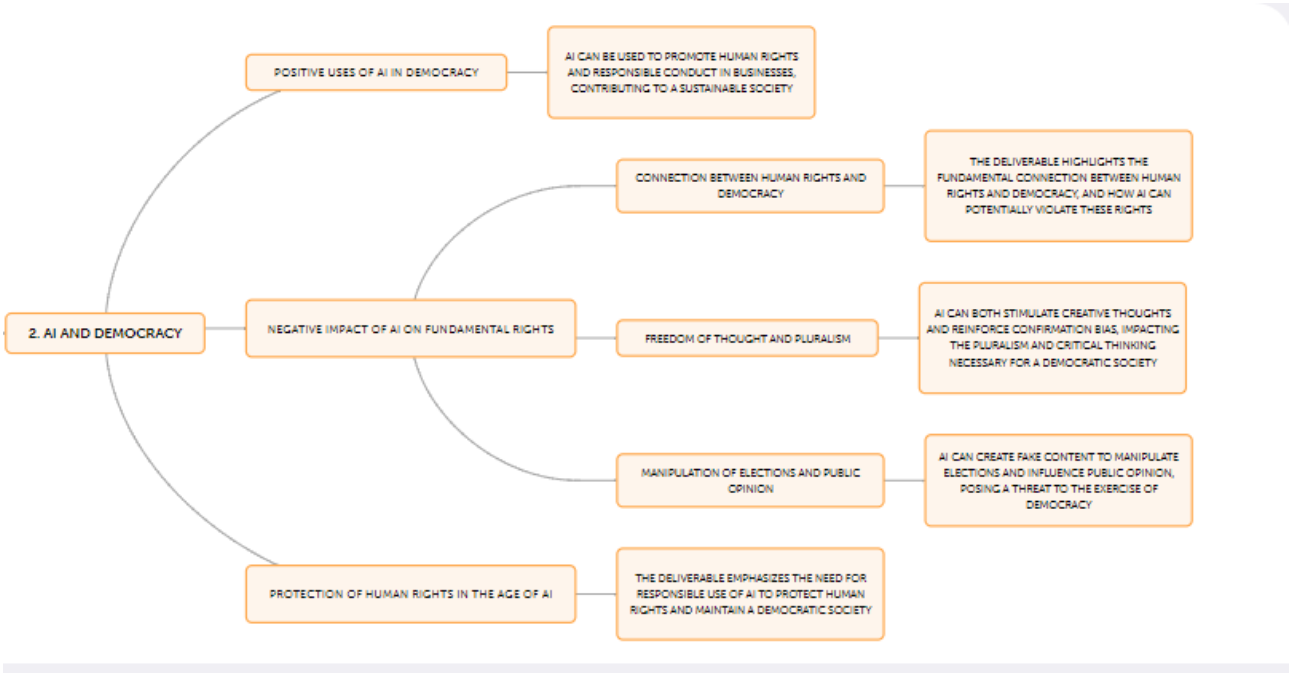

Figure 6. AI and Democracy

2.3 AI and elections

One of the greatest concerns about AI and Generative Adversial Networks (GANs) is related to elections. This is particularly important in democracies, where free and fair elections are the basis upon which the system is built and maintained. Democracy relies on electoral accountability. Voters are asked to elect candidates they support or to remove candidates from office which don't have any more the electoral support which brought them there. Therefore, democracy requires a healthy information environment where voters can monitor what politicians are doing, learn what candidates are promising to do when elected, and assess which policies might be needed in response to societal challenges.

Generative AI has the potential to disrupt the voters information environment in two ways. First, generative AI allows for the creation of highly convincing deepfakes-–images and videos that are difficult for non-specialists to distinguish from genuine content. Second, AI chatbots are a new, direct source of information for certain voters, especially younger ones.

Deepfakes

The dissemination of deepfakes during an electoral campaign, has increased considerably in the last years. JUST TWO DAYS before Slovakia's elections, an audio recording was posted to Facebook. On it were two voices: allegedly, Michal Šimečka, who leads the liberal Progressive Slovakia party, and Monika Tódová from the daily newspaper Denník N. They appeared to be discussing how to rig the election, partly by buying votes from the country's marginalized Roma minority. Šimečka and Denník N immediately denounced the audio as fake. The fact-checking department of news agency AFP said the audio showed signs of being manipulated using AI. But the recording was posted during a 48-hour moratorium ahead of the polls opening, during which media outlets and politicians are supposed to stay silent. That meant, under Slovakia's election rules, the post was difficult to widely debunk.

Deepfakes played a role in the recent Turkish elections. At a large political rally, President Recep Tayyip Erdoğan showed a fake video linking his chief opponent, Kemal Kılıçdaroğlu, to the leader of the PKK, a Kurdish group classified by the State Department's as a foreign terrorist organization. Separately, an online Kılıçdaroğlu supporter used AI to generate a video that appeared to show the candidate delivering a campaign speech in perfect English.

During electoral campaigns deepfakes create false content, with the purpose of manipulating voters. These could have the effect of making a voter change sides at the last minute, before he has the time necessary to verify if the video or the audio is real or fake, or it could have the other effect of making a voter not vote at all, because of the uncertainty that a video or audio may create. It is particularly concerning that AI-manufactured content could be released very close to election day in order to generate fake scandals within a time frame that makes fact checking difficult.

Second, wide-spread circulation of manufactured content may undermine voters' trust in the broader information environment. If voters come to believe that they cannot trust any digital evidence, it becomes difficult to seriously evaluate those who seek to represent them. Third, politicians may use this undermining of the credibility of the information environment to dismiss genuine information. Late in the Turkish election, a tape came to light showing compromising images of a candidate, Muharrem İnce. While İnce eventually withdrew, he also claimed the video was a deepfake. If voters genuinely can't tell the difference between what is fake and what is real, it is not hard to imagine that such denials will become a commonplace.

Microtargeting and manipulation

In 2016, Cambridge Analytica, a firm specializing in using online data to create voter personality profiles in order to target them with messages, became famous swaying voters from one political candidate to another based on exploitation of voters' particular psychological vulnerabilities. This tactic involves deducing psychological attributes that are not readily observable, such as personality traits, from individuals' online behavior and personal data. Subsequently, these inferred psychological features are leveraged to craft highly personalized messages tailored to each individual. Cambridge Analytica was involved in the Vote Leave campaign (United Kingdom), the 2016 Trump campaign (United States), and other political campaigns spanning 68 countries before it folded in 2018 after investigations opened in several countries. The investigation by a British Parliamentary committee concluded that relentless targeting that plays "to the fears and the prejudices of people, in order to alter their voting plans" is "more invasive than obviously false information" and contributes to a "democratic crisis". Microtargeting is also problematic outside the political domain when it exploits people's moment-to-moment emotional state. Facebook has access to technology that can identify vulnerable teenagers at moments when they feel "worthless" and "insecure," although the technology was ostensibly never made available to advertisers and only used in an experimental context.

Findings indicate that political microtargeting is an effective technique and can be automated using off-the-self generative AI. In this case, the generative AI could help improve microtargeting, fashioning political ads better customed towards each person targeted. Generative AI could improve the attempt to use vast amounts of online data to establish individuals' personality traits and use this information to create remarkably persuasive political campaigns.

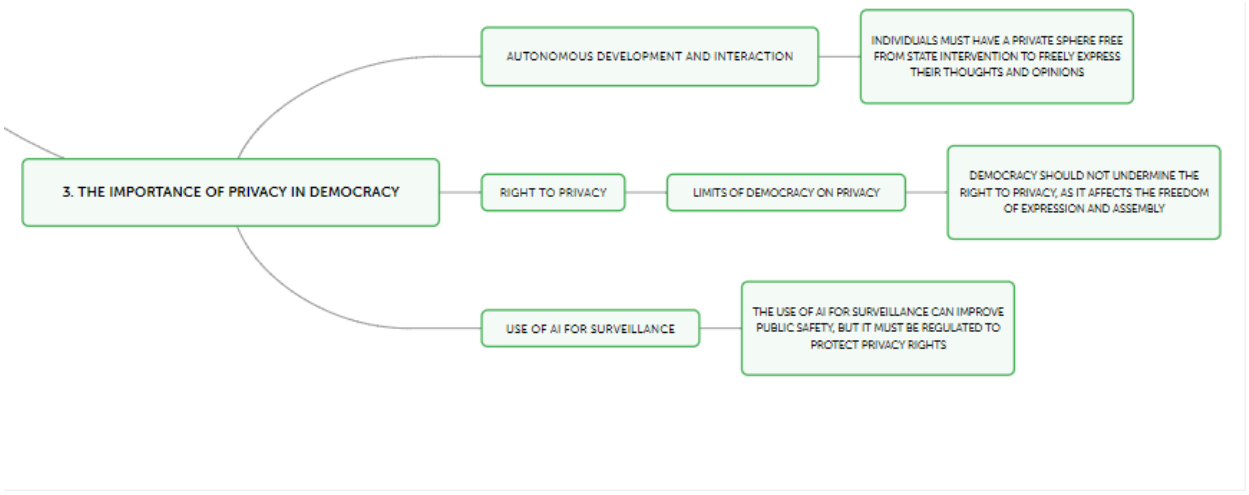

Figure 7. The importance of privacy in democracy

How effective these technics really are?

The question is if these technics are really effective and can they help change the outcome of an election. Until now, there hasn't been any case when a campaign supported by Generative AI has influenced so many elections that has helped change the outcome. As generative AI is used more and more to create fake videos, audios, images, texts, etc., voters also are more knowledgeable about this phenomenon and are not easily manipulated by them, or tend to believe them easily and the further these phenomena are discussed in public, the more difficult will be for them to manipulate or persuade voters.

Studies on political persuasion show that political microtargeting's persuasive returns may in fact be limited. But in close elections, a small proportion of extra voters can make all the difference. This is of particular concern in democracies such as the US, where the results in closely divided swing states can decide the Presidential election by just a few thousand votes.

3 The geopolitical implications and risks associated with GANs diffusion

3.1 Detailed Conceptual Analysis

Geopolitical risks are defined as the potential political, economic, military, and social risks that can emerge from a nation's involvement in international affairs. Typically, they emerge whenever there is a major shift in power, a conflict, or a crisis. These risks can have far-reaching implications for both the country itself and the global community at large. There are many factors that can contribute to geopolitical risks, such as a nation's economic stability, its political relations with other countries, and its military strength. In recent years, globalization has also played a role in exacerbating these risks by increasing the interconnectedness of the world's economies and societies.

Technological Dualism and Political Influence

Bernard Fallery's exploration in "Regards critiques sur l'Intelligence Artificielle, les intérêts politiques des empires numériques" presents a critical view of the dichotomy in the political influence of AI and GANs. This dualism echoes through the corridors of power in the Western Balkans, where the tug of war between democratic ideals and authoritarian tendencies is palpable. Fallery articulates how North American tech giants (GAFAMI) champion a libertarian approach that emphasises individual autonomy and self-control, contrasting sharply with the state-centric control model pursued by Chinese tech conglomerates (BHATX), which underscores a more authoritarian, control-oriented approach. This distinction is not merely theoretical but has practical implications for how information technologies are deployed and regulated within the geopolitical fabric of the Western Balkans, reflecting broader global power dynamics (Fallery, 2021).

Cybersecurity and Geopolitical Risks

In "Impacts and Risk of Generative AI Technology on Cyber Defense," Subash Neupane and colleagues provide a comprehensive overview of the multifaceted risks that GANs and AI pose to cybersecurity and, by extension, to geopolitical stability. The paper outlines how these technologies can be exploited for malicious purposes, including public opinion manipulation, election interference, and the facilitation of cyberattacks. The authors argue that the ability of AI and GANs to enhance reconnaissance capabilities and evade detection poses a significant challenge to national and international security architectures. This analysis is particularly relevant for the Western Balkans, a region characterized by its strategic geopolitical position and history of conflict, making it susceptible to cyber threats and informational warfare (Neupane *et al.*, 2023).

Deepfake Dilemma and Public Perception

The document "GANs Gone Wild: Public Perceptions of Deepfake Technologies on YouTube" sheds light on the alarming rise of deepfake technologies. These AI-generated falsifications are not only a technical marvel but also a significant threat to the integrity of information and public trust. The ability of deepfakes to convincingly replicate real individuals saying or doing things they never did has profound implications for political discourse and the media landscape. For countries in the Western Balkans, where societal trust is already fragile and political landscapes are volatile, the potential for deepfakes to further erode trust in democratic institutions and exacerbate social divisions is a pressing concern. The paper underscores the need for comprehensive policy and legislative frameworks to address the challenges posed by deepfakes, highlighting their geopolitical implications.

AI's Role in Geopolitical Dynamics

Nicolas Miailhe's "Géopolitique de l'Intelligence artificielle: le retour des empires" offers a macroscopic view of how AI and GANs are reshaping the contours of global geopolitics. Miailhe posits that the rapid advancement of these technologies is central to the emergence of digital empires, predominantly led by the United States and China. The document outlines how AI influences military capabilities, economic strategies, and political power, highlighting the strategic importance of AI development in global competition. This narrative is especially pertinent to the Western Balkans, where external geopolitical pressures and internal aspirations for European integration intersect, making the region a focal point for digital and geopolitical contestation (Miailhe, 2018).

Media Integrity and Information Warfare

KATI BREMME's "MédIAs, Nouvelle génération" delves into the transformative impact of GANs and AI on media and journalism. BREMME discusses the challenges posed by the rapid evolution of these technologies,

including the spread of misinformation, the ethical dilemmas surrounding content creation, and the erosion of traditional media models. The document highlights how the blurring line between truth and fiction, propelled by AI's capability to generate realistic content, raises significant geopolitical concerns regarding information reliability, mass manipulation, and the overall integrity of democratic processes. In the Western Balkans, where the media landscape is often polarized and susceptible to political influence, these challenges are magnified, affecting public perception and trust in an already complex geopolitical context (Bremme, 2023).

The 5-Year Spam and Chinese Influence Operations

"The 5-year Spam: Tracking a Persistent Chinese Influence Operation" provides an insightful analysis into Dragonbridge, a sophisticated influence campaign aligned with Chinese government interests. This document illustrates the multifaceted nature of modern influence operations, employing a diverse toolkit that includes social media manipulation, multimedia content creation, and strategic narrative construction. The operation's impact on global geopolitics, including efforts to shape perceptions of China and its policies, underscores the complex interplay between technology, information warfare, and geopolitical strategy. For the Western Balkans, where external influences vie for power and influence, understanding the mechanics and implications of such operations is crucial for safeguarding national integrity and democratic values.

Organized Chaos and the Balkans' Fake News Ecosystem

"Organized Chaos" specifically addresses the pervasive issue of fake news in the Western Balkans, providing a granular analysis of its manifestations and impacts across the region. By examining specific instances, such as Albania's governmental response to misinformation, the document highlights the nuanced challenges posed by disinformation campaigns. These campaigns not only distort public discourse but also exploit historical grievances and political divisions, complicating the region's path toward democratic consolidation and European integration. The analysis presented in "Organised Chaos" is indispensable for understanding the broader geopolitical and social ramifications of fake news in the Western Balkans, offering insights into the strategies and countermeasures needed to combat misinformation.

Conclusion

The detailed exploration of the geopolitical implications and risks associated with GANs, fake news, and AI in the Western Balkans reveals a complex tapestry of technological innovation, political maneuvering, and societal impact. As the region navigates these challenges, the insights from the aforementioned documents provide a valuable framework for understanding and addressing the multifaceted nature of digital technologies in geopolitical contexts.

Another very relevant aspect related to the impact of generative artificial intelligence on international geopolitical arrangements is the so-called 'Hi-Tech colonisation': this is a form of digital colonialism where the use of digital technologies is aimed at the political, economic and social domination of another nation or territory. While with classical colonialism, Western nations seized foreign lands and appropriated indigenous knowledge to incorporate it into industrial processes, with the advent of digital colonialism, the spread of digital technologies and artificial intelligence has become deeply integrated with the conventional tools of capitalism and authoritarian governance, such as labour exploitation, policy capture, economic planning, secret services, ruling class hegemony and propaganda.

3.2 The Hi-Tech colonization

Another very relevant aspect related to the impact of generative artificial intelligence on international geopolitical arrangements is the so-called 'Hi-Tech colonisation': this is a form of digital colonialism where the use of digital technologies is aimed at the political, economic and social domination of another nation or territory. While with classical colonialism, Western nations seized foreign lands and appropriated indigenous knowledge to incorporate it into industrial processes, with the advent of digital colonialism, the spread of digital technologies and artificial intelligence has become deeply integrated with the conventional tools of capitalism and authoritarian governance, such as labour exploitation, policy capture, economic planning, secret services, ruling class hegemony and propaganda.

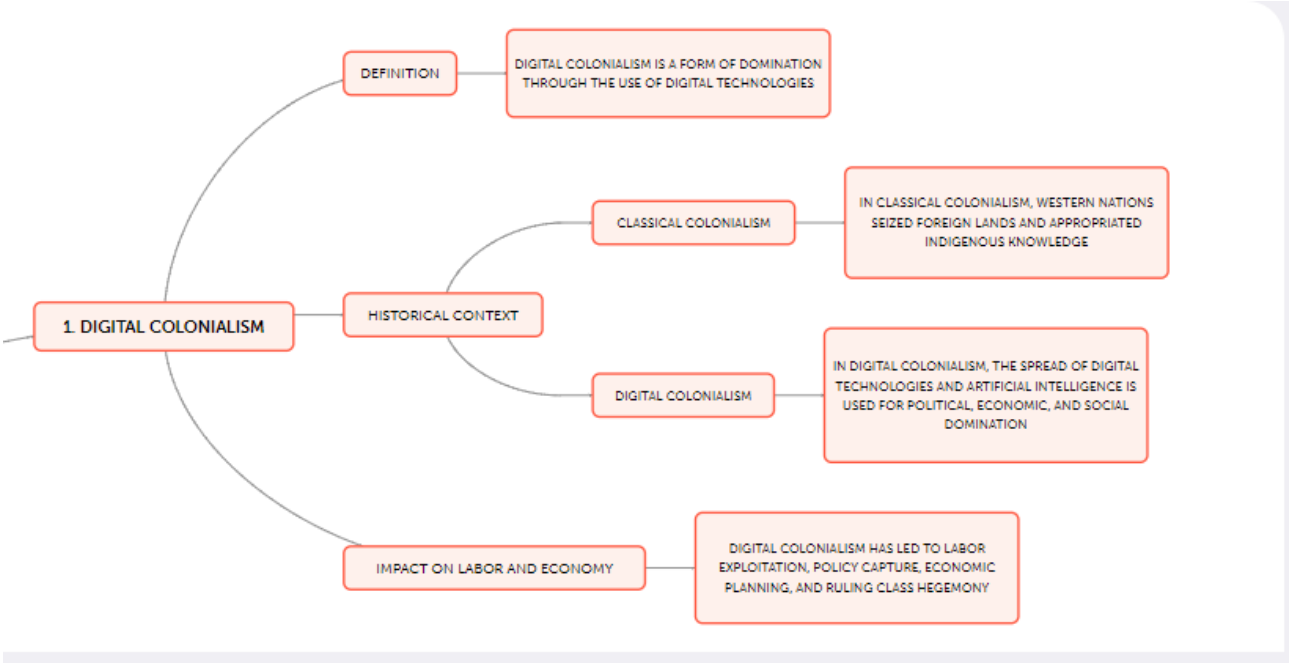

Figure 8. Digital colonialism

The privatisation of software has been accompanied by the rapid centralisation of the Internet in the hands of intermediary service providers such as Facebook and Google. Essentially, this shift to cloud services has nullified the freedoms that FOSS licences guaranteed to users, because software is run from the computers of Big Tech multinationals. Corporate clouds expropriate people from the ability to control their own computers. Cloud services provide petabytes of information to corporations, which use the data to train their artificial intelligence systems. Artificial intelligence uses Big Data to 'learn' - it needs millions of images to recognise, for example, the letter 'A' in different fonts and formats. Applying this to humans, the sensitive data of people's private lives becomes a resource of incalculable value that technology giants relentlessly try to extract. In other words, the technology giants control the business relationships throughout the production chain, profiting from their knowledge, accumulated capital and hegemony of key functional components.

Digital colonialism, therefore, becomes a barycentric practice for the consolidation of an unequal division of labour, in which the dominant powers use ownership of the digital infrastructure, scientific knowledge and control of the means of calculation to keep certain nations in a situation of permanent dependence. However, this unequal division of labour has evolved. Economically, production has moved down the value hierarchy, replaced by an advanced hi-tech economy in which Big Tech companies are firmly in command. Digital colonialism is rooted in the domain of 'things' in the digital world that form the means of computation: software, hardware and network connectivity. It includes the platforms that act as gatekeepers, the data extracted by intermediary service providers and industry standards, and the private ownership of 'intellectual property' and 'digital intelligence'.

The contest is very complex. The ecological crisis created by capitalism is seriously threatening to destroy life on earth, and solutions for a digital economy must intersect with environmental justice and a broader battle for social equality. To eliminate the phenomenon of digital colonialism, we need an ethical paradigm capable of questioning the purely economic ends of hi-tech imperialism in order to put the human being at the centre of its universal value.

This means, as Paolo Benanti reminds us, to initiate an ethical transition within the use of digital technologies and artificial intelligence in particular so that the data collected does not serve to increase the economic gain of a few technocratic elites but is directed towards the promotion of human flourishing that can involve everyone.

The consequences of digital colonialism in education should also be borne in mind here: it is spreading rapidly in the educational systems of many countries. Schools are good sites for Big Tech to expand control of the digital market. In those countries where governments provide students with a device at no cost, multinationals are able to acquire and capitalise on a significant amount of data. This is a way to retain the new generation in the use of specific software and platforms. In doing so, however, not only do students become true guinea pigs from which to obtain data, but they are also oriented towards the use of specific platforms that will most likely

be preferred by them in the future. In the face of these scenarios, a reflection is required on the need to curb digital colonialism in order to foster a fair and transparent distribution of digital resources with the sole aim of fostering the human in its fullness, beyond any social or economic discrimination. It is therefore necessary to rethink digital technologies no longer as a private service for the benefit of a few, but as a public good that requires to be disciplined and regulated in its use in order to ensure maximum transparency, social justice, and avoid the exploitation of the weakest categories of the world's population.

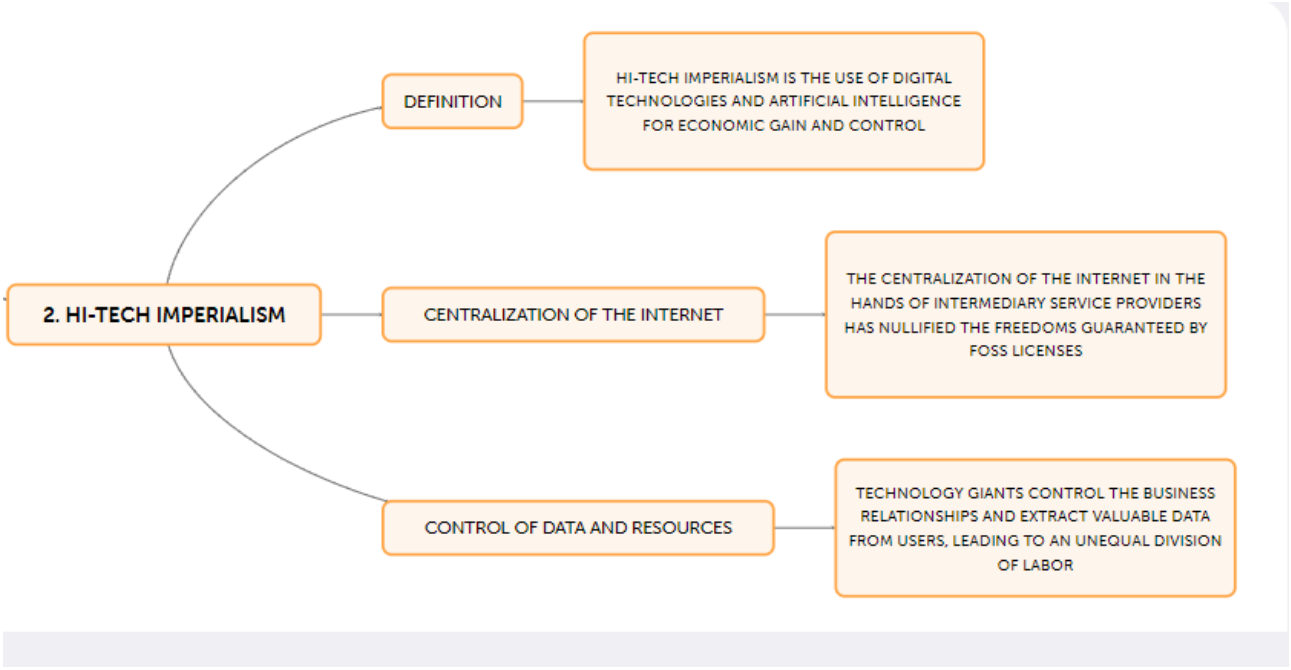

Figure 9. Hi-Tech imperialism

4 Comparative Law frameworks

In this part, an attempt is made to work out a methodology for comparative legal research, which goes beyond the 'functional method' or methodological scepticism. Comparative research is still mainly about comparing national legal systems, even if different forms of globalization, such as Europeanization, and an increasing recognition of non-state law, such as customary law, religious law or unofficial law-making by international companies are challenging the very concept of 'legal system'.

4.1 A comparative look

This project will also explore the regulation and in general terms the legislative frameworks that are being proposed and implemented in the field of AI and in particular in the field of Generative AI in different, key countries of the world. In particular, it will study a number of cases:

- The United States
- China
- India
- UK

A cursory look at the previous cases yields the following results:

a) United States

A very good summary of the current situation in the United States at the Federal level is given by this article, published in the New York Times in July 21, 2023 (Kang, 2023). The summary of the summary would be that regulating AI is very much under discussion in the US, a country where no specific regulation whatsoever has been yet adopted in this field. The article also points at the fact that the EU is behind the European Union when it comes to AI regulation. Apparently, the philosophy in the United States is that regulating AI could severely hinder the development of a very profitable market. The United States is the global champion in this market, in competition with the Chinese. The risk of "going for private commitments" and self-regulation of the main economic players in this market is very much in place, despite the fact that some of these key players (Sam Altman, the Chief Executive of OpenAI, in particular) have asked to the Congress to be regulated.

A different question is the States' level in the United States. Here progress is more important than the one that has been achieved at the Federal level. According to this Brooking Institution policy brief (Engler, Friedler and Venkatasubramanian, 2023), the State that has make more headway in this terrain has been California, which has adopted the California Artificial Intelligence Accountability Act in September 2022. Other States have adopted or updated more sectoral legislation in this field.

b) China

China is possibly one of the countries of the world that has taken regulation of AI (and regulation of generative AI) more seriously. In this policy paper of the Carnegie Endowment for International Peace (Sheehan, 2023), one can find a very good summary of what China is doing at this regard. The Chinese regulatory approach to AI seems to be more inductive and bottom-up than deductive and top down. With this we mean that the Chinese authorities have adopted a number of sectoral administrative regulations (for example, in the field of Generative AI) to then check how they work in practice. After this sequence of particular and sectoral administrative regulations, it is now drafting a general law on AI. The different pieces of this regulatory framework are tied up together through a crucial document: the 2017 New Generation AI Development Plan (see Roberts et al., 2021, for a summary of this plan). The AIDP focuses on three major aspects: AI innovation for military uses; AI as a boost of economic growth; and AI social and above all ethical governance. The Chinese Government has a very clear understanding of a "first-mover" advantage in the regulatory field. It is

trying to set the stage in the ongoing global debate on AI regulation, and therefore, it is trying to anticipate itself to regulation coming from the EU and the US.

c) India

India is probably the country of the world that has adopted the most aggressive strategy regarding the deployment of AI. At present, there is no specific regulatory framework in the field of AI in India. The Indian Commission NITI Aayog (a governmental agency entrusted with think-tank tasks) has produced a number of reports ((2020, 2022; 2021) on the matter, in which it basically calls for the setting up of a regulatory framework in the field of AI. The public debate on regulation of AI was prompted in this country when the IT Minister, Ashwini Vaishnaw, declared, in a written response to the Lok Sabha -the Indian Parliament- (April 2023), that the Indian government was not planning to regulate AI (2023). However, after a visit of Sam Altman to Prime Minister Modi in June 2023, the State IT Minister (Secretary of State), Rajeev Chandrasekhar declared that the Indian government would regulate AI "to keep digital citizens safe" (2023). Therefore, whether AI will be regulated or not, and if the first option is retained, the extent to which AI will be regulated, is a discussion very much in flux right now in this country.

d) The UK

As we know, the UK is no longer a member of the European Union, since February 1, 2020. One of the sectors in which we can corroborate the direct impact that UK's exit from the Union has produced, is precisely the field of AI. The UK approach to AI regulation clearly differs from the one that has been adopted by the EU. The UK government published in March 2023 its White Paper on AI (2023). Building trust in AI is cited amongst the main aims of the UK's regulatory approach to AI. In this White Paper, the UK government makes it clear that it opts for a principled approach to regulation of AI. This principled approach means that the government will only issue principles that will be addressed to the different UK agencies that deal directly or indirectly with AI. These principles are: safety, security and robustness; transparency and explainability; fairness; accountability and governance; contestability and redress. UK agencies will, in turn, adopt regulatory standards for each of the sectors that fall under their respective sphere of competence. Therefore, the UK's approach is non-statutory and sector-by-sector. Norms in place will be administrative norms. The White Paper says that it aims, with this approach, to give flexibility to the field of AI. Instead of a top-down, regulatory (legislative) approach, as the one that has been adopted in the EU, the UK has opted for a sectoral approach in which regulatory (administrative) standards will be adopted on the basis of the specificities of each sector and the principles that will be issued by the UK's government on governance of AI.

e) Conclusion

Our cursory look at the approach to AI regulation of key countries of the world yields one clear conclusion: there is no common approach to regulation of AI in the world. The extremes of the segment are formed, on the one hand, by the EU, that has opted for a rules (legislative) based and top-down, general regulation of AI, and on the other hand, by countries like the UK and probably India, which have opted for much more flexibility in this field. The state of the art in regulation of AI is very much in flux: it will therefore evolve in the future, as we know more about the actual dangers and risks that are prompted by the use of AI. The regulation of, in particular, generative AI, will therefore depend of the general approach to regulation of AI that each of these countries or economic areas will adopt.

This has clear consequences for our project. As we are in the EU, we are somehow constrained by the regulatory approach that the EU institutions have adopted in this field. The idea is, as has been mentioned before, that a general AI regulation will be in place in the close future. Therefore, any proposal for regulation of generative AI must start from this regulatory constriction. However, it could be possible, from there, to adopt more flexible and principled regulations regarding generative AI.

In general, the dilemma that regulators are confronted with in the field of AI, both in general terms and in particular, in the sector of Generative AI, is clear. Regulators would like to inject a certain degree of legal certainty in this field without stifling the development of what is a very promising market. This equilibrium is balanced more on the side of the market in some cases (India and the UK) and more on the side of legal certainty and security in some others (the EU and probably China). The US would represent a middle ground in this context, at least for now. But at the end of the day, it is clear that market considerations will be predominant in this field. This is why opting for a much more flexible, discretionary and regulatory (administrative) approach in this field might be an interesting second-best to regulating the field of AI and generative AI. It will then all depend on the hubris of the regulatory agencies that are set in place in this field.

References

https://www.nytimes.com/2023/07/21/technology/ai-united-states-regulation.html

https://www.brookings.edu/articles/how-california-and-other-states-are-tackling-ai-legislation/

https://fagenwasanni.com/news/the-regulation-of-artificial-intelligence-at-the-state-level-in-the-united-states/75877/

https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117

https://niti.gov.in/sites/default/files/2020-07/Responsible-AI.pdf

https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf

https://www.niti.gov.in/sites/default/files/2022-11/Ai_for_All_2022_02112022_0.pdf

https://economictimes.indiatimes.com/tech/technology/not-considering-any-laws-to-regulate-ai-growth-in-india-it-minister-ashwini-vaishnaw/articleshow/99275493.cms

https://timesofindia.indiatimes.com/india/we-will-regulate-artificial-intelligence-and-any-emerging-technology-based-on-user-harm-it-minister-rajeev-chandrasekhar/articleshow/100873366.cms

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1146542/a_pro-innovation_approach_to_AI_regulation.pdf

Conclusion

This Deliverable, in accordance with T4.1 of the SOLARIS project, aimed to present an interdisciplinary mapping of political risks and implications related to infodemics generated by GANs. The transversal and multidisciplinary method allowed to enucleate the most relevant aspects of the use of generative artificial intelligence in the context of international geopolitical arrangements by interweaving theoretical paradigms with punctual analyses of phenomena.

First, the political actors with their respective interests involved in the spread of GANs were presented and the connection between the circulation of certain specific content with particular interests of a geopolitical nature was shown (a); then, the impact of such technologies in political communities where the exercise of democratic freedom risks being compromised by infodemic manipulation and the circulation of dystopian information aimed at affirming authoritarian systems of thought was discussed (b). Thus, an analysis of the geopolitical implications of the spread of generative AI allowed us to ascertain the impact of this technology in the areas of international cybersecurity, information integrity and the constitution of new informational ecosystems (c). A brief reference to the phenomenon of digital colonialism made it possible to highlight how hi-tec power concentrated in the hands of a few multinational agencies runs the risk of availing itself of forms of social exploitation both vis-à-vis less developed nations and the new generations, as is the case, for instance, with educational digital colonialism (d). Finally, a comparative look at the international level of the laws enacted to regulate these technologies offered an articulate and very effective picture of the state of the art of the main contemporary legal systems as regards the regulation of Gans (e).

What is offered in this Deliverable constitutes the outcome of a shared and transversal research path whose aim is to interrogate current phenomena in order to put them at the service of present and future humanity. For this reason, the theoretical gains of this research are aimed at questioning a manipulative use of these technologies aimed at the spread of infodemics and the assertion of authoritarian powers, in order to promote a humanisation and ethical circulation that knows how to use these tools in a fair and democratic manner, for the good of all human beings involved in the current digital revolution.

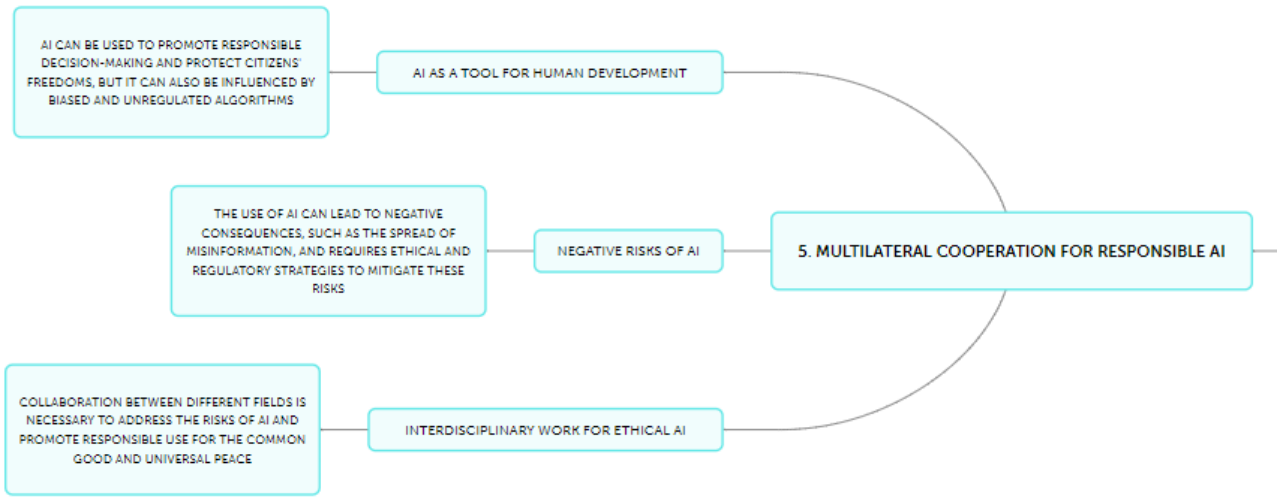

Figure 10. Multilateral cooperation for responsible AI

References

A pro-innovation approach to AI regulation (2023). London: Dandy Booksellers Ltd.

Aizenberg, E. and Van Den Hoven, J. (2020) 'Designing for human rights in AI', *Big Data & Society*, 7(2), p. 205395172094956. Available at: https://doi.org/10.1177/2053951720949566.

Alesina, A. and La Ferrara, E. (2000) 'Participation in Heterogeneous Communities*', *Quarterly Journal of Economics*, 115(3), pp. 847–904. Available at: https://doi.org/10.1162/003355300554935.

Approach Document for India Part 1 – Principles for Responsible AI (2021). NITI Aayog. Available at: https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf.

Baptista, E. (2022) 'Insight: China uses AI software to improve its surveillance capabilities', *Reuters*, 8 April. Available at: https://www.reuters.com/world/china/china-uses-ai-software-improve-its-surveillance-capabilities-2022-04-08/ (Accessed: 16 May 2024).

Bareis, J. and Katzenbach, C. (2022) 'Talking AI into Being: The Narratives and Imaginaries of National AI Strategies and Their Performative Politics', *Science, Technology, & Human Values*, 47(5), pp. 855–881. Available at: https://doi.org/10.1177/01622439211030007.

Bender, E.M. *et al.* (2021) 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜', in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. *FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency*, Virtual Event Canada: ACM, pp. 610–623. Available at: https://doi.org/10.1145/3442188.3445922.

Bradford, A. (2023) 'The Race to Regulate Artificial Intelligence', *Foreign Affairs*, 27 June. Available at: https://www.foreignaffairs.com/united-states/race-regulate-artificial-intelligence (Accessed: 16 May 2024).

Bremme, K. (2023) 'MédIAs Nouvelle génération', *PRINTEMPS-ÉTÉ 2023* [Preprint]. Available at: https://www.meta-media.fr/wp-content/uploads/sites/33/2023/07/metamedia-22-pages.pdf.

Buchanan, B. *et al.* (2021) *Truth, Lies, and Automation: How Language Models Could Change Disinformation*. Center for Security and Emerging Technology. Available at: https://doi.org/10.51593/2021CA003.

Bunz, M. and Braghieri, M. (2022) 'The AI doctor will see you now: assessing the framing of AI in news coverage', *AI & SOCIETY*, 37(1), pp. 9–22. Available at: https://doi.org/10.1007/s00146-021-01145-9.

Chun, Wendy Hui Kyong (2024) *Discriminating Data: correlation, neighborhoods, and the new politics of recognition*. S.l.: MIT PRESS.

Coeckelbergh, M. (2020) *AI ethics*. Cambridge, MA: The MIT Press (The MIT press essential knowledge series).

Coeckelbergh, M. (2022) 'Democracy, epistemic agency, and AI: political epistemology in times of artificial intelligence', *AI and Ethics* [Preprint]. Available at: https://doi.org/10.1007/s43681-022-00239-4.

Crawford, K. (2021) *Atlas of AI: power, politics, and the planetary costs of artificial intelligence*. New Haven London: Yale University Press.

Engler, A., Friedler, S. and Venkatasubramanian, S. (2023) 'How California and other states are tackling AI legislation', *Brookings*, 22 March. Available at: articles/how-california-and-other-states-are-tackling-ai-legislation/.

Fallery, B. (2021) 'Regards critiques sur l'Intelligence Artificielle, les intérêts politiques des empires numériques', *HAL Open Science* [Preprint]. Available at: https://hal.science/hal-03126059/document.

Fyler, T. (2023) 'X goes back to the future on content moderation', *T_HQ*, 30 August. Available at: https://techhq.com/2023/08/whats-behind-x-hiring-content-moderation-staff-and-allows-political-advertising-again/#:~:text=X%20announces%20new%20content%20moderation%20hiring%20ahead%20of%202024%20election.&text=The%20platform%20is%20also%20allowing,the%20first%20time%20since%202019.

Helmus (2022) *Artificial Intelligence, Deepfakes, and Disinformation: A Primer*. RAND Corporation. Available at: https://doi.org/10.7249/PEA1043-1.

Indian Commission NITI Aayog (2020) *Responsible AI*. Indian Commission NITI Aayog. Available at: https://niti.gov.in/sites/default/files/2020-07/Responsible-AI.pdf.

Indian Commission NITI Aayog (2022) *Responsible AI*. Indian Commission NITI Aayog. Available at: https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf.

Jones, M.O. (2022) *Digital authoritarianism in the Middle East: deception, disinformation and social media*. London: Hurst & Company.

Kang (2023) 'In U.S., Regulating A.I. Is in Its "Early Days"', *The New York Times*, 21 July. Available at: https://www.nytimes.com/2023/07/21/technology/ai-united-states-regulation.html.

Katz, Y. (2020) *Artificial whiteness: politics and ideology in artificial intelligence*. New York: Columbia University Press.

Kwet, M. (2019) 'Digital Colonialism: South Africa's Education Transformation in the Shadow of Silicon Valley', *SSRN Electronic Journal* [Preprint]. Available at: https://doi.org/10.2139/ssrn.3496049.

Lynch, J. (2022) 'Iron net: Digital repression in the Middle East and North Africa', *European Council on Foreign Relations*, 29 June. Available at: https://ecfr.eu/publication/iron-net-digital-repression-in-the-middle-east-and-north-africa/ (Accessed: 16 May 2024).

Marwick, A. and Lewis, R. (2017) *Data & Society Media Manipulation and Disinformation Online Case Studies*. Data & Society Research Institute. Available at: https://datasociety.net/output/media-manipulation-and-disinfo-online.

McGuffie, K. and Newhouse, A. (2020) 'The Radicalization Risks of GPT-3 and Advanced Neural Language Models'. Available at: https://doi.org/10.48550/ARXIV.2009.06807.

Miailhe, N. (2018) 'Géopolitique de l'Intelligence artificielle : le retour des empires ?:', *Politique étrangère*, Automne(3), pp. 105–117. Available at: https://doi.org/10.3917/pe.183.0105.

Natale, S. (2021) *Deceitful media: artificial intelligence and social life after the Turing test*. New York, NY: Oxford University Press.

Neupane, S. *et al.* (2023) 'Impacts and Risk of Generative AI Technology on Cyber Defense'. Available at: https://doi.org/10.48550/ARXIV.2306.13033.

Novelli, C. *et al.* (2024) 'Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity', *SSRN Electronic Journal* [Preprint]. Available at: https://doi.org/10.2139/ssrn.4694565.

Roberts, H. *et al.* (2021) 'The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation', *AI & SOCIETY*, 36(1), pp. 59–77. Available at: https://doi.org/10.1007/s00146-020-00992-2.

Sheehan, M. (2023) 'China's AI Regulations and How They Get Made'. Carnegie Endowment for International Peace. Available at: https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117.

The Economic Times (2023) 'Not planning any law to regulate AI growth in India: IT minister Ashwini Vaishnaw', 5 April. Available at: https://economictimes.indiatimes.com/tech/technology/not-considering-any-laws-to-regulate-ai-growth-in-india-it-minister-ashwini-vaishnaw/articleshow/99275493.cms.

Times of India (2023) 'Govt will regulate AI to keep digital citizens safe; tech poses no risk to jobs in next 5 years: Union IT minister Rajeev Chandrasekhar', 9 June. Available at: https://timesofindia.indiatimes.com/india/we-will-regulate-artificial-intelligence-and-any-emerging-technology-based-on-user-harm-it-minister-rajeev-chandrasekhar/articleshow/100873366.cms.

Umbrello, S. and Van De Poel, I. (2021) 'Mapping value sensitive design onto AI for social good principles', *AI and Ethics*, 1(3), pp. 283–296. Available at: https://doi.org/10.1007/s43681-021-00038-3.

Waivv, P. (2021) 'GANs Gone Wild: Public Perceptions of Deepfake Technologies on YouTube'. Available at: https://dodo.is.cuni.cz/handle/20.500.11956/150489?show=full.

Warren, P. *et al.* (2023) 'The 5-year Spam: Tracking a Persistent Chinese Influence Operation'. Available at: https://tigerprints.clemson.edu/mfh_ci_reports/7.

Weikmann, T. and Lecheler, S. (2023) 'Cutting through the Hype: Understanding the Implications of Deepfakes for the Fact-Checking Actor-Network', *Digital Journalism*, pp. 1–18. Available at: https://doi.org/10.1080/21670811.2023.2194665.

Xiang, N. (2023) 'China's AI sector has no time for end-of-the-world worries', *Nikkei Asia*, 18 December. Available at: https://asia.nikkei.com/Opinion/China-s-AI-sector-has-no-time-for-end-of-the-world-worries.