PARLIAMENTARY ASSEMBLY OF THE MEDITERRANEAN
ASSEMBLEE PARLEMENTAIRE DE LA MEDITERRANEE
برلمان البحر الأبيض المتوسط

Notes on "AI and Parliament: Ensuring the Integrity of our Digital Democracy" Conference

Commonwealth Parliamentary Assembly,

4-6 December 2023, online

Executive summary:

From December 4th to 6th, 2023, the Parliamentary Assembly of the Mediterranean (PAM) actively participated in a virtual conference titled "AI and Parliament: Ensuring the Integrity of our Digital Democracy", organized by the Commonwealth Parliamentary Assembly (CPA). Hon. Ray Abela, a Member of the Parliament of Malta and PAM's rapporteur on digitalization, represented PAM with an intervention centered on deepfakes and the protection of human rights.

The primary focus of the conference laid in the intricate relationship between Artificial Intelligence (AI) and parliamentary systems, with a pivotal emphasis on AI's transformative potential to enhance lives and societies globally, contingent upon the safeguarding of individuals' rights. Discussions underscored the dual nature of AI, serving as a force for positive change and a substantial threat.

Over the course of the three-day conference over 20 AI experts offered interventions that delved into critical concerns and challenges posed by AI. The initial day provided a comprehensive overview of the core concepts related to AI, disinformation and synthetic media. The second day was dedicated to an exploration of the main threats and vulnerabilities regarding disinformation, along with strategies to address these challenges. The concluding day focused on individual protection, specifically examining the potential harms that deepfakes and disinformation may cause to the personal reputation of key political figures within the electoral domain.

Throughout the conference, it became increasingly clear that parliamentarians shall fully grasp the multifaceted threats posed by AI and synthetic media to create a robust legislative framework fostering accountability, transparency, and authenticity in the realm of AI. This requires a thorough acknowledgment of the risks associated with AI, coupled with a commitment to set the basis for a legislative environment that addresses these risks effectively.

The conference served as an invaluable platform to gain an understanding of the latest developments in the field of AI. It allowed participants to engage with experts and colleagues from around the world. By fostering the exchange of knowledge and collaborative dialogue, the event played a substantial role in empowering parliamentarians to navigate the landscape of AI, ultimately contributing to the advancement of responsible and forward-looking governance.

Day 1 - Core Concepts: Disinformation, Artificial Intelligence and Synthetic Media

Session 1 – Opening remarks

Stephen Twigg, *CPA Secretary-General*

- The SG Twigg underlined the regular engagement of CPA with parliaments and legislators on a set of crucial issues related to technology and innovation
- The conference aims to provide an accessible space where people can exchange ideas and opinions
- The use and the risks of AI in national elections are particularly crucial. 2024 will be a crucial year for parliaments, as elections will be held in the U.S., India, Pakistan, and several countries in Africa and Europe

Session 2 – A History of Disinformation in Democracy

Dr. Heidi Tworek, *Canada Research Chair, Director, Centre for the Study of Democratic Institutions, Associate Professor, University of British Columbia*

- Misinformation and disinformation are two different concepts. **Misinformation refers to false or inaccurate information** (getting the facts wrong), while **disinformation refers to the deliberate spread of false information** by malign actors, with harmful aims, often spurred by underpinning political or economic goals.
- A key concern regards the reasons why certain people or states engage in creating and spreading false information

Hon. Chloe Smith, *UK MP, former UK Secretary of State on Innovation and Technology*

- AI encompasses a wide range of different fields, such as education, business, healthcare, defense and the standards of living.
- If the right calls are made and fair principles followed, AI may really give rise to a revolution favoring equality, responsiveness, and fairness. This may be true in the electoral field as well.
    - i.e., Accelerating AI investments, deployment and capabilities may represent an enormous opportunity for the public good, as it could foster economic growth (generating up to 7 trillion dollars of revenues in the next ten years).
- **However, there are also several risks that come with AI. The lack of a common positive opinion about the use of AI implies the need for people to be reassured about the potential risks of AI**.
- National governments may have different strategies. For example, the British government maintains a preventive approach. The British "Frontier AI Task Force" is the first state organization that focuses on advancing AI for public safety, strengthening security in different fields ranging from crime prevention to election' monitoring

Session 3 – Generative AI and Synthetic Media: Present and Future Perspectives

Dr. Carolyn Ashurst, *Senior Research Associate in Safe and Ethical AI at the Alan Turing Institute*

- The use of language models in technology is broad, as it ranges from analyzing a text and answering questions, to styling a piece of writing. However, language models retain limitations. For example, differently from human beings, these models are unable to judge and understand the trustworthiness of a text as they take into account all data they have available indiscriminately. This inability to differentiate between data generates concerns regarding the trustworthiness of these resources

Cassidy Bereskin, *Director & Founder, OxGen AI, and author of the forthcoming CPA Handbook on Disinformation, AI and Synthetic Media*

- Synthetic media, also referred to as generative media, is defined as visual, auditory, or multimodal content that has been generated or modified (commonly via artificial intelligence). Such outputs are highly realistic, as they simulate artifacts, people and events, and they cannot be identified as synthetic to the average person
- These synthetic media outcomes may have an impact on the economic and social sphere of society.
    - i.e., The picture portraying an explosion next to the Pentagon caused a fall in the stock market[1]
- There are two clear threats that arise from synthetic media
    - 1. They may damage a person's reputation
    - 2. They may affect the fairness and transparency of a democratic process, undermining citizens' trust and undermining the credibility of reliable sources

Tommy Shaffer Shane, *AI policy advisor at the Center for Long-Term Resilience*

- There is a **need for standards, ethical norms, and transparency in big AI companies as well as public regulators at the international level.**
- It is necessary to improve the control mechanisms for AI and data mining. Detection tools to identify deepfakes will improve but will be always based on probability and percentage, leaving a margin for error
- Moreover, it is crucial to ensure equal access to security technologies in order to combat the use of deepfakes globally

[1] Haddad, M. (2023). Fake Pentagon explosion photo goes viral: How to spot an AI image. Retrieved from www.aljazeera.com website: https://www.aljazeera.com/news/2023/5/23/fake-pentagon-explosion-photo-goes-viral-how-to-spot-an-ai-image

Session 4 – Disinformation, Deepfakes and Elections

Veronika Hincová Frankovská, *Project Manager at Demagog.SK*

- In September 2023, Slovakia held parliamentary elections and there was a great focus on the War in Ukraine
- A general fear regarding the possible electoral fraud spread among citizens.
  - AI-generated audio recordings started to circulate on the Internet. For example, an audio in which one candidate promised to increase the price of beer once elected was circulated one night prior to the opening of the electoral polls, affecting the electoral results
  - Another example is represented by an AI-generated audio circulating on Telegram and Facebook, in which the Slovakian journalist, Monika Tódová, states that the electoral result was manipulated. According to Globsec and CEDMO, over 50% of the Slovakian population believed the electoral outcome to be manipulated after hearing that audio.

Silas Jonathan, *Head of Digital investigation and OSIMT at the Center of journalism and development at Dubawa*

- Unlike Western elections, where disinformation often revolves around policy and political party differences, the 2023 Nigerian Presidential elections were marked by distinctive challenges, such as religious and ethnocentric disinformation, coordinated disinformation in different chambers, attacks targeting political candidates…
- Several Facebook groups concurrently propagated misleading content on social media platforms, magnifying the impact of disinformation during the Nigerian elections
- In the Nigerian electoral context, fact-checkers play a pivotal role in transcending fact-finding. Indeed, the aim of fact-checkers should be to identify bad electoral actors, with a focus on key political figures. AI is a double-edged sword for fact-checkers in Nigerian elections. On the one hand, they can represent a useful tool to strengthen fact-checking. On the other hand, AI tools can be employed to spread disinformation spreading deepfakes, fake electoral results before the official closure of polls, and creating chaos undermining the democratic process

Day 2: "The Threat Landscape: Threats, Vulnerabilities and Combatting Disinformation"

Mr. Brody McDonald, *Associate Fellow at the International Centre for the Study of Radicalisation (ICSR)*

- Mr. McDonald delves into the intricate landscape of actors engaging with AI tools, emphasizing the nuanced variations in sophistication and objectives. This spectrum spans from individuals casually experimenting with consumer-level tools to extremist and terrorist groups leveraging advanced AI for propaganda and disinformation. What becomes evident is the transformative impact of AI on the tactics and strategies of malicious actors
- The evolving nature of **extremism is particularly concerning**, as it is marked by a shift towards decentralized structures and the prominence of *lone wolf* actors. In this regard, Mr. McDonald underscores the alarming empowerment of smaller, malicious entities facilitated by the consumerization of AI tools. **Unlike in the past, where extensive resources were required, these tools now enable a single individual**, perhaps situated in a basement, to wield influence on a scale previously unimaginable
- A critical highlight is the **escalating utilization of AI tools by state actors in the realm of disinformation**. According to Mr. McDonald, the prospect of industrial-scale operations by countries like **Iran, Russia, or North Korea** poses a considerable challenge. Notably, the sophistication of state-driven campaigns surpasses that of individuals or extremist groups, leveraging vast resources, IP hiding, and a relative lack of government crackdown
- Mr. McDonald delivered specific examples from the last six to eight months, illustrating the real-world consequences of AI-driven disinformation. The ongoing situation in **Gaza**, marked by a proliferation of disinformation on various platforms, serves as a poignant illustration. Additionally, incidents such as protests in the UK, where AI-driven disinformation fuelled far-right groups, demonstrate the tangible impact on public events and sentiments
- The erosion of public trust emerges as a significant concern, with Mr. McDonald discussing the **"spoilers dividend". This phenomenon involves exploiting public uncertainty by denying or attributing shocking statements to AI-generated content, further muddling the already complex landscape of truth and falsehood**
- Mr. McDonald advocated for a pragmatic framework acknowledging that **disinformation is an inevitable byproduct of technological progress**. Drawing parallels to environmental pollution, Mr. McDonald suggested that should focus on **containment and mitigation**, rather than complete eradication. This perspective prompts a discussion on the necessary interventions to limit the adverse effects of disinformation
- Mr. McDonald presented a two-fold approach, addressing both the developer and distribution sides of AI-driven disinformation. **On the developer side, Mr McDonald stressed the importance of provenance tools, notably blockchain technology, to provide a transparent trail of an image's creation and manipulation.** While acknowledging the contributions of industry leaders like Adobe and Microsoft, there are concerns about the risks posed by smaller startups that lack resources to implement robust trust and safety measures
- Turning to the distribution side, Mr. McDonald identified the **critical role played by alternative tech platforms,** such as Gab and Truth Social, **in the viral dissemination**

of disinformation. This underscores the necessity of extending regulatory measures beyond mainstream platforms to encompass these emerging players.

- Finally, Mr. McDonald advocated for a **balanced regulatory approach that avoids impinging on free speech.** Indeed, the proposed interventions should not jeopardize the principles of free expression and free speech, which are fundamental to the democratic values the international community seeks to protect.

Q&A Session

In the ever-evolving landscape of AI-generated disinformation, where do we stand in terms of development and usage by state and non-state actors? Is the global society witnessing a maturity in the logic and sophistication of these tools, moving beyond the early experimental stages? Moreover, can people assert that AI-generated disinformation has transitioned from being a supplementary tactic to becoming a primary tool for state actors and other entities?

Mr. Brody McDonald

- Mr. McDonald shed light on the varying degrees of sophistication exhibited by different actors. F**oreign state governments, equipped with substantial resources, engage in advanced and sophisticated disinformation campaigns**. This echoes the cascading quality of operations, where well-funded entities establish robust mechanisms for deploying AI tools in disinformation.
- The experimental use of AI by **extremist groups** is not yet widespread. **There is a swift evolution in this field**, citing examples of groups in Pakistan employing AI-generated images to enhance the appeal of audio recordings. The integration of contextual **translation tools** further amplifies the impact, allowing these groups to convey messages in 240 languages with remarkable accuracy, in real-time synchronization with video content.
- With the layering of different AI tools technologies are combined to broaden the reach and enhance the engagement of disinformation content. This strategic approach involves incorporating contextual translation tools alongside other AI-generated elements, creating a more immersive and captivating disinformation narrative.
- The discussion touched upon **individual actors experimenting with AI tools**, exemplifying a case where a malicious actor circulated supposed photographs of Palestinians engaging in festivities around the world during the Hamas-Israel conflict. Despite being swiftly addressed on platforms like Twitter and Telegram, this incident highlighted the evolving nature of disinformation tactics at the hands of individual actors. Mr. McDonald underscored the challenges faced by the community tasked with tackling unsophisticated content, emphasizing that the leaps in sophistication over the last six months are significant.
- A notable concern regards the potential convergence of state-level and individual-level tools in terms of power and sophistication. He cautioned that, **given the rapid advancements witnessed, the once easily identifiable and amateurish nature of individual tools may soon rival those employed by state actors**. Looking ahead, this trajectory raises serious apprehensions about the efficacy of current mechanisms to identify and counter disinformation, necessitating a proactive and adaptive approach to address the evolving landscape of digital threats.

Hon. **Marco Mendicino**, *House of Commons, Canada*

- Hon. Mendicino commended the CPA for choosing "Artificial Intelligence, Disinformation, Parliament, and Digital Democracy" as the conference theme, emphasizing the timely nature of this topic. Acknowledging **AI's potential to add over $4 trillion to the global economy annually by 2030**, hon. Mendicino discussed the transformative power of AI in addressing critical issues such as climate change, healthcare, and poverty.
- Turning attention to AI's impact on democracy, hon. Mendicino explored the positive applications, such as informed voter decision-making and interactive communication between leaders and constituents. However, he also raised concerns about the potential misuse of AI, citing examples of deepfakes and their rapid dissemination of misinformation.
- Highlighting the **UNESCO report identifying threats to elections**[2], hon. Mendicino stressed the need to mitigate risks while leveraging AI for the benefit of democratic institutions. The focus of the discussion shifted to legislative and policy reforms, with hon. Mendicino proposing **a new paradigm based on generally accepted AI principles.** These principles encompass **safety, transparency, accountability, privacy, equity, openness, digital literacy, and responsibility for the public good**.
- Hon. Mendicino then delved into the current state of democracy, noting a decline in the number of democracies and public confidence. The presentation shifted to the second part, discussing where society needs to go. This involved a detailed examination of each principle and its application to the electoral process, emphasizing the paramount importance of safety and transparency.
- Expanding on safety, hon. Mendicino explored the **dual responsibility of developers and governments**, stating that **the focus is on preventing AI misuse in political campaigns** through reasonable safety parameters, including the prohibition of using AI to depict opponents violating the law. Hon. Mendicino suggested leveraging government agencies to monitor and report serious incidents, promoting lawful and responsible AI use.
- Moving to transparency, hon. Mendicino underscored its significance in maintaining trust within the democratic process. Concerns raised about AI's potential to create deceptive content, challenging traditional transparency measures. Hon. Mendicino emphasized the need for **higher ethical standards**, especially in global campaigns, to combat the evolving capabilities of generative AI.
- Hon. Mendicino outlined a roadmap for the future, emphasizing three broad recommendations for aligning the modernization of democratic institutions with the evolution of AI. The final plea is **a call to action, urging legislators to collaborate with developers, consult the public, and enact new laws and policies to create a framework for responsible AI use in democracy**.
- Hon. Mendicino thereby underscored the need for increased transparency in the digital era of political campaigns, emphasizing openness about data, algorithms, automated decision-making, and AI-generated outputs.
- Addressing the potential misuse of generative AI in political advertising, hon. Mendicino suggested restrictions and even prohibitions for portraying opponents engaging in illegal or violent acts. Balancing free speech against regulatory measures,

[2] Achler, M., Krimmer, R., Kužel, R., Licht, N., Rabitsch, A., & UNESCO. (2022). Elections in digital times. UNESCO Publishing.

he acknowledged the nuanced challenges, especially in cases where AI is used to portray opponents negatively

- Moving to accountability, hon. Mendicino emphasized the crucial role of human oversight in elections, asserting that humans must retain the right to oversee compliance and the final vote. Acknowledging cautionary notes from developers, he called for self-accountability within the AI sector to promote responsible AI development
- Discussing privacy, hon. Mendicino reflected on the complex consumer dilemma of agreeing to terms without thorough scrutiny. He then highlighted privacy commissioner recommendations, calling for strengthened language recognizing privacy as a fundamental right and ensuring flexible dispute resolution
- The principle of equity is also emphasized, urging AI development that reduces systemic barriers, promotes inclusivity, and avoids perpetuating historical biases. In the context of elections, ensuring equity is vital to prevent the malicious use of AI that could impede accessibility or suppress votes.
- Hon. Mendicino addressed digital literacy, stressing the need for **public education campaigns** to enhance understanding of AI in elections. He eventually proposed the integration of digital literacy into education, emphasizing the importance of critical reasoning to combat misinformation
- The final principle of responsibility in the public good was discussed, advocating for AI development that benefits society as a whole. Aligning with global challenges, **responsible AI should balance innovation with societal welfare to avoid harm and misuse**
- Looking ahead, hon. Mendicino endorsed concrete **suggestions**, including an **International Convention to establish global standards for AI**. The urgency of acting swiftly and decisively was emphasized, acknowledging the evolving nature of AI and its potential impact on democracy

Mr. **Andy Parson**, *Senior Director of Engineering at Adobe* and **Mr**. **Bruce MacCormack**, *Neural Transform and Project Origin*;

- The role of **provenance** is crucial in **addressing issues of misinformation and preserving democratic values and human rights**. In fact, **innovative initiatives are underway to combat the subversive nature of AI-generated disinformation, especially during elections and sensitive political periods worldwide**
- Mr. Parson and Mr. MacCormack emphasized the **need for a shift towards proving the truth rather than merely detecting misinformation**. They discussed challenges with detection methods and the ethical implications of detectors in creating feedback loops. They underscored the **importance of trust and verifiability in content, stressing the need for cryptographic assurance**
- The discussion touched upon the **Coalition for Content Provenance and Authenticity (C2PA), a nonprofit entity developing technical standards**. Mr. Parson highlighted their commitment to privacy-preserving measures, allowing individuals to maintain anonymity if they choose not to expose specific information
- Mr. Parson and Mr. MacCormack also addressed the challenges of AI-generated information, acknowledging its potential misuse by various actors, both state and non-state. They questioned the current level of AI development and its effectiveness, especially in differentiating between authentic and deceptive content
- The discussion delved into whether AI-generated information has reached a level of maturity where it could be exploited as a primary tool by malicious actors. Concerns

were raised about the implications for security and whether the logic governing AI models had reached a sufficiently advanced stage to counter potential threats effectively

- Moreover, they pondered over the trade-off between providing full transparency and maintaining the comprehensibility of the information being presented. The conversation questioned who holds the responsibility for building literacy around these standards, whether it is at the individual level, or at the governance level
- In addressing these complex issues, Mr. Parson emphasized the importance of understanding the user context. He differentiated between **the general public's consumption of information and that of investigative reporters, highlighting the need for different levels of scrutiny**. The discussion touched upon the **challenges faced by social media companies in managing the overwhelming volume of information and the role of AI in aiding fact-checking processes**.
- There is a need for a balanced approach able to ensure that trust signals and disclosure mechanisms are culturally sensitive and adapted to various contexts and languages. Both speakers emphasized that AI can be a valuable tool in creative and storytelling pursuits when used responsibly and with transparency. The overarching goal is to provide users with the necessary information to make informed judgments without instilling unnecessary fear about synthetic media.
- In the ongoing conversation, Mr. Parson and Mr. MacCormack not only acknowledged the challenges surrounding AI-generated information but also delved into the broader implications for media, especially during critical events like elections. They considered **the pressure on media outlets to deliver instant coverage amid the demand for breaking news, particularly in conflict zones where individuals on the ground use smartphones to share immediate and unverified content**
- The discussion raised pertinent questions about the potential tension between the need for swift reporting and the imperative to ensure the accuracy and trustworthiness of the information being disseminated. They explored the dynamic landscape where competitors strive to release news quickly, and the challenges this presents for media outlets in verifying the authenticity of AI-generated content
- Mr. Parson brought attention to the journalistic art of fact-checking, emphasizing that AI tools, while useful for validation, are just one aspect of a journalist's toolkit. He drew distinctions between scenarios where media outlets have established relationships with contributors and instances where content arrives from unknown sources, highlighting the varying levels of trust associated with each
- Moreover, the conversation touched on the content supply chain, discussing the value of maintaining provenance even as media leaves reputable sources and circulates on social media platforms. This aspect brought attention to the **complexities faced by media outlets in ensuring authenticity throughout the distribution process**
- As the discussion shifted towards the urgency of addressing these challenges, Mr. Parson and Mr. MacCormack expressed the need for **solutions like cryptographic signatures to establish the provenance of media**. They acknowledged the regret that efforts to establish trust in media did not commence earlier and highlighted the potential of deploying solutions like cryptographic signatures to combat misinformation

Concluding remarks

- In the final segment, the conversation issued a call to action for legislators, policymakers, and regulators. Mr. Parson and Mr. MacCormack emphasized the role that standards, particularly the **Content Authenticity Initiative (CAI)**, could play in

addressing the challenges posed by AI-generated content. They urged policymakers to **support initiatives focused on media trust and authenticity, underscoring the necessity for collaborative efforts across stakeholders to ensure the integrity of information in an increasingly digital landscape**.

Day 3 – "Individual Protection: Deepfakes, Disinformation and Reputational Threats"

Session 1 – A general overview

Professor. Marietje Schaake, *Standford University Professor*

- In the discussion regarding the extent to which Gen AI can be used to manipulate news, it is important to take a balance between acknowledging the existence of fake news and the necessity to consider the news reliable and avoid labeling everything as fake
- According to Professor Schaake, one of the main problems regarding **Gen AI is that it is created by private companies**. However, there is a clear mismatch of incentives between the private and the public sector. These companies have **incentives to push their products in the market faster than their competitors**
- **The decision-making power of the Parliamentarians is stolen by big tech companies that create AI and provide cybersecurity and news**. Consequently, and inevitably, big techs play a role in the decision-making process of these crucial fields
- Moreover, big tech companies play a role in the **geopolitical sphere** such as in the conflicts of Gaza and Ukraine
- The monopoly of power and responsibility that used to lie in the hands of the State, is now falling into the hands of the private actors
- According to Professor Schaake, **the main problem of AI regards governance**. It is not the misuse of AI in elections *per se*, but rather a much more systemic problem that concerns all aspects of the governance of society[3]

Hon. Ray Abela, *PAM Rapporteur on Digitalization (Malta)*

- Deep fakes and synthetic media pose a threat to Parliamentarians and public leaders
- Gen AI is often employed to create content that cannot be distinguished from reality
    - i.e., the UN SG Guterres was presented with a video of himself speaking perfectly in Chinese. He had the surreal experience of seeing himself delivering a speech in fluent Chinese, despite being unable to speak Chinese. This is an example of the potential dangers of AI today.
- The impact of Gen AI in the decision-making process in several areas is increasing. Examples encompass the healthcare system, such as health monitoring.
- AI can also impact politicians and infringe on the fairness and transparency of the electoral processes. Synthetic media can be used in disinformation campaigns and erode public trust in public institutions. Public awareness is important, and people need to receive a suitable education regarding AI and need to protect themselves:
    - *Authentication and verification* – promote robust verification, so that people can distinguish between genuine and manufactured information

[3] University, S., Stanford, & California 94305. (n.d.). Marietje Schaake. Retrieved December 13, 2023, from cyber.fsi.stanford.edu website: https://cyber.fsi.stanford.edu/people/marietje-schaake

- *Cybersecurit*y – being cautious about sharing data, changing passwords and being aware of fake identities
- *Human rights protection* – enhance a legislative framework that can create regulation for content creation, distribution and punishment for malicious use

Professor Alexander Evans, *OBE Professor at the London School of Economics (LSE)*

- When talking about synthetic media, some people believe this is just the evolution of political satire and commentaries
- Some of it has a long tradition in political cartoons and satire.
  - i.e., sometimes the cartoon version of a politician is better known to the population than the politician himself
- AI and ICTs can be used to frame and present the same material in different ways
- Gen AI creates content, such as deepfakes, that are often deployed to mine the legitimacy of a candidate in the electoral campaign
- Synthetic speech and voice recording are harder than pictures to be debunked
- **The greatest damage to candidates is infringed whenever the deepfakes are dropped just before the closure of polls so that it is too late for the candidate to correct the damage**

Hon. Ray Abela

- Marketing and information have evolved. The manipulation of information has existed since the paper era. However, with social media information travels faster

Professor Alexander Evans

- In the past, political parties and governments have tended to be fast followers rather than innovators. The fact that they will do the same with AI is not worrying
- It is key to have digital forensic capabilities and the ability to hold someone accountable. Even private actors must be held accountable

Hon. Ray Abela

- Digital forensics are needed to see whether deepfake material is uploaded to the web

Session 2 – Synthetic media and human rights: Building a Rights-Based Approach

Loui Mainga, *Program Communications Coordinator, Africa, WITNESS*

- According to M. Mainga, the spread of fake news risks undermining information, above all when it comes from minority groups. The inability to tell what is real and what is fake favors the spread of fake news
  - i.e., a video showing police brutality against a minority group can be discredited as false, due to the decreased public trust in videos circulating online.

Authorities carrying out brutalities easily discredit the sources and the disseminated material, ending up spreading false information

- Other examples have been seen in the war in Ukraine or elections in Africa, with fake news fomenting violence among tribes, chaos and violence, and discrediting politicians
- Three guiding principles regulating AI should be:
    - Protect human rights
    - Placing responsibility across AI pipelines
    - Create critical infrastructure, laws, and regulations
- Local communities need skills, tools, and capacities to detect manipulation. Detection equity is also an issue. The question is not just to grant access to those tools, but also to provide the skills to use these tools

Professor Nnenna Ifeanyi-Ajufo, *Professor of Law and Technology, Leeds Law School, Leeds Beckett University*

- The concern about human rights violations is that they happen even when AI is deployed for good reasons and used to protect humans. This issue raises the **question of whether human rights violations are inherent to the use of AI**
- **AI risks generating systemic marginalization, creating new inequalities and amplifying old ones**
- Tech companies are the leaders and, due to their powerful nature, the dissemination of AI will involve little discussion of accountability (regarding human rights issues)
- Moreover, it is necessary to assess the reality in the African region, with regard to financial availability and racial discrimination. AI risks exacerbating racial discrimination, while the low financial availability risks hindering the African development of AI compared to the rest of the world
- Design technology is also about including and focusing on the audience and the insights you have from people. Big techs need to come up with creative solutions and test the technology for feedback before deploying it, and consider privacy, freedom of speech, racial discrimination ... The **approach must be human-centric**
- However, tech companies claim to use the human-centered approach, but there is no actual definition
- AI calls for a nature of obligation and human rights must be embedded around decision-making, legislation and policies

Loui Mainga

- Transparency in AI is important for human rights. However, 60% of the population has social media in the world, and there are slightly under 400 million users in Africa. It is hard to expect that all people will have a suitable skill set to identify the threats and deepfakes.
- **Middle literacy capabilities are desirable, and this would be a great antidote, but it is unsustainable. People should strengthen transparency instead**

Professor Nnenna Ifeanyi-Ajufo

- When thinking about cybersecurity, it is necessary to think about culture, participation and literacy. People must be able to understand the value of the technologies that are deployed. Transparency is key, but it is part of literacy

Session 3 – Deepfakes: a new online Gender-Based Violence (GBV)?

Suzie Dunn, *Senior Fellow at CIGI, a Ph.D. candidate at the University of Ottawa and an Assistant Professor of Law & Technology at Dalhousie University, Canada*

- According to Ms. Dunn, deepfake is an overused term. Synthetic media includes deepfakes but is any visual material where AI has been used to manipulate and create the images. Differently, deepfakes require a base video that takes the face of another person and over-imposes another face
- Facial re-enactment is when the face moves and it is not superimposed
- Voice cloning only regards the sound and it is harder to detect than deep-fakes
- Deepfakes can create different harms:
    - Imaged-based sexual abuse;
    - Identity attacks;
    - Political interference (even though there are actually not many examples of political deepfakes that are successful in impacting politics)
- **Image-based sexual abuse** is an umbrella term that covers a variety of sexual abuses. This is a form of GBV. Women are overrepresented in the number of incidents compared to men and the impact on women is much higher because of the existing stereotypes
- 96% of deep fakes are non-consensual sexual deep fakes of women
- Initially the victims were mainly celebrities in the U.S., UK and South Korea, but now there are several deepfakes of politicians, particularly made during elections
- Everyday people increasingly show up in deepfakes as well
- Harms that deepfakes can cause encompass stress impacts and mental health impacts, such as limitations to bodily autonomy and integrity issues. Moreover, deepfakes silence women because the fear that their images could be used to produce deepfakes disincentives women to play an active role online
- The ex-post reaction to harms caused by deepfakes is present, but there is a lack of preventive work that should be implemented by the government, like education and standing up against deepfake abuse

Kiran Hassan, *coordinator of freedom of expression and digital rights at the Institute of Commonwealth Studies*

- The hyper-sexualization of female Parliamentarians has harshly increased with the advent of social media
- For this reason, young females are less inclined to join politics and run for public office
- Despite having several laws and policies, there is a lack of a comprehensive social and legal framework. The crisis for information is not for the legal community to act alone. There is a need for a proactive response from the governments and civil groups as well.

There should be a multilateral conversation where the government cooperates with civil society and the private sector
- Tech companies should be held accountable. In this respect, the multilateral governmental conversation becomes very important

Varaidzo Faith Magodo-Matimba, *Grants and Growth coordinator at "Pollicy*"
- Pollicy advocates for a better digital future, highlight issues and advocate for solution for women
- The violence experienced by women online hinders their participation in politics and society and often incentivizes their withdrawal from the public scene.
- The Internet should be tailored to the specific needs of women in certain regions, such as African women

Concluding remarks

Sophie Compton, *Director, Producer, and Activist of "Another Body" documentary*
- *"Another Body"* is a documentary that talks about a young girl who experiences deep fake and sees herself acting in pornography. The case was closed because no law was broken in the State of the US where the case was carried out
- From the documentary, it emerges clearly that no woman in the world is safe from this technology
- Women who undergo this kind of experience often experience issues of trust and anxiety thereafter and are prone to withdraw from their social circle fearing that someone could see the video
- Current society has reached the point where everyone acknowledges that every celebrity probably has a deep fake online. But this cannot and should not be normalized
- Victim blaming is also very common for deepfakes
- Government and parliamentarians need to tackle the normalization of this abuse and consider the landscape where this abuse is occurring, in order to meaningfully address it

MC/AG/GR/0712/2023
Updated: GR/1312/2023