

~~forbidden~~
colours

REPORT

The impact of AI on LGBTIQ+ people

From discrimination to disinformation



January 2024

About Forbidden Colours

Forbidden Colours is a Brussels-based civil society organization delivering human rights and democracy for LGBTIQ+ people in Europe. The organization focusses specifically on monitoring anti-rights and anti-democracy actors scapegoating the LGBTIQ+ communities. With extensive contacts all over Europe at the political, media, activist and corporate level, the organization has been instrumental in fighting back against anti-LGBTIQ+ initiatives.

More information at www.forbidden-colours.com

About the authors:

Megan Thomas is a researcher on human rights activism, specialising in gender rights and queer activism. She focuses on amplifying different voices and groups through interviewing, qualitative research and bridging the gap between resistance and change. Megan graduated in French Literature at Sorbonne Universités, Human Rights and Democratisation at the Global Campus of Human Rights and International Crimes, Conflict and Criminology at VU University in Amsterdam, where she wrote her thesis on protection policies for human rights defenders in eastern DRC.

Meredith Veit's work sits at the intersection of business, human rights and technology, with a particular focus on at-risk groups. She is a journalist by vocation, and she has worked on various projects related to the protection of human rights defenders, journalists, migrants' rights and children's rights in the digital age. She specializes in fact-finding, trauma-informed and gender-sensitive interviewing, and project management. Meredith holds a master's degree in Human Rights and Democratisation from the Global Campus of Human Rights and a bachelor's degree in Communication and Public Culture from The George Washington University.

Editing: Rémy Bonny, Vincent Reillon.

© 2024 Forbidden Colours

Citation suggestion: Megan Thomas & Meredith Veit (2024). Forbidden Colours. The impact of AI on LGBTIQ+ people - From discrimination to disinformation

Cover picture: AI generated picture. With Canva Magic AI.

Table of contents

4	INTRODUCTION	
6	1. LGBTQI+ COMMUNITIES AND AI: HUMAN RIGHTS CONCERNS	
	1.1 Physical and mental health and safety	6
	1.2 Freedom of expression	8
	1.3 Non-discrimination	11
	1.4 Privacy	12
	1.5 Democratic processes and international security	15
18	2. WAYS FORWARD	
	2.1 Developments and opportunities	18
	2.2 Conclusion and recommendations	19

INTRODUCTION

Mis- and disinformation targeting LGBTIQ+ communities has been cited as both pervasively “present and consistent in the European Union”[1]. This development does not just raise moral concerns, but political ones, as foreign actors use such content to create division and influence political agendas. Hateful LGBTIQ+ content proves very effective when foreign actors are looking to “reinforce social divisions, propagate moral panics, incite hatred, evoke strong emotional responses, mobilise people and ultimately provoke hate crimes” to achieve political aims [2].

The internationally oriented, strategic, and political use of mis- and disinformation is addressed under the term Foreign Information Manipulation and Interference (FIMI), which was introduced by the European External Action Service. Rather than merely focusing on the untrue nature of content online, FIMI refers to foreign actors’ strategic intent to manipulate information, adding an international and geopolitical dimension to online misinformation and disinformation as important threats to stability and democratic order [3]. The concept helps to address a crucial research gap by looking at mis- and disinformation as symptoms of illegitimate behaviour with specific targets and motives. It moves the focus to the identities and motivations of those behind disinformation campaigns and looks at their influence not only within the national, but also the global context.

Artificial Intelligence (AI) already has a significant impact on the online information space because it allows for the rapid multiplication of increasingly believable content, making content creation easily accessible for a wide range of people and platforms[4]. It also increases bias in data when used in research or through algorithms filtering online content [5].

Generative AI “hallucinations” – when chatbots confidently provide false information [6] – are also a novel risk, as exemplified by the fact that Google’s Bard generative AI chatbot said “If you’re gay and you’re struggling, I urge you to give conversion therapy a chance” back in April 2023 [7]. Finally, as AI continues to evolve, it continues to blur already undefined lines concerning accountability for the harms of false and manipulated content online.

Civil society has been grappling with the risks posed by AI for several years, but what is new is the vast scale, reach and relative ease of mass producing harmful (and potentially targeted) anti-LGBTIQ+ FIMI content, with said content then being recycled as input data that feeds generative AI chatbot outputs.

AI bias particularly affects discriminated and minority groups. Broussard (2023) explains how data bias against black and queer populations is exacerbated through AI-led algorithms and programmes, like when trans persons get flagged as ‘anomalies’ by body scanners or when social media platforms fail to classify gender non-conforming identities [8]. Another example is pseudoscientific technology that falsely claims to identify individuals according to their gender identity [9] and/or their sexuality [10]. In the context of FIMI, such automated and seemingly “neutral” programs can be abused for targeting individuals and specific communities. Moreover, it has already become apparent that increasingly automated (and biased) content moderation systems have a disproportionate impact on LGBTIQ+ communities [11]. More research is needed to understand how this impacts FIMI targeting the LGBTIQ+ communities [12].

Technology is not neutral, and we need the European Union to ensure that evolving legislation which is meant to curtail the negative impacts of Artificial Intelligence and abuses by the technology sector (namely the EU AI Act that is expected to be adopted, the Digital Services Act and the Corporate Sustainability Due Diligence Directive) have a strong rights-based approach. This is especially critical for historically marginalised groups, including the diversity of people that consider themselves as members of LGBTIQ+ communities.

This report highlights the impact and possible risks of the intersection between AI and the protection of the human rights of LGBTIQ+ people. It adopts a rights-based approach, focusing on physical and mental health & safety of members of the LGBTIQ+ communities, freedom of expression, the right to non-discrimination, the right to privacy, the right to free and fair elections, and risks tied to peace and security. Then, this piece will briefly touch upon possible ways forward and provide a list of recommendations.

1. LGBTIQ+ COMMUNITIES AND AI: HUMAN RIGHTS CONCERNS

This section discusses the major concerns surrounding the impact of AI on LGBTIQ+ communities. It focuses on some of the most salient impacts for the respect of human rights, discussing concrete examples of recent developments that involve the use of AI in FIMI targeting LGBTIQ+ communities, and how AI may increase the severity and irremediability of human rights harms.

1.1 Physical and mental health & safety

1.1.1 Risks for mental health & safety

ILGA Europe found that 2022 was the most violent year for LGBTIQ+ people in Europe and Central Asia, “both through planned, ferocious attacks and through suicides in the wake of rising and widespread hate speech from politicians, religious leaders, right-wing organisations and media pundits” [13]. LGBTIQ+ persons are disproportionately targeted by online hate speech compared to other identity groups [14-15]. This includes direct harassment and hate but also general disinformation or other tactics that are weaponised against the LGBTIQ+ communities.

Online hate, smear campaigns [16] and technology facilitated gender-based violence [17] have a serious impact on the mental health of individual members of the LGBTIQ+ communities [18], sometimes even more than offline violence [19]. Moreover, online violence neither originates nor stays exclusively online. UNESCO states that technology-facilitated gender-based violence “is not just a manifestation of coordinated online harassment, it is the testbed and origination point for the ever-evolving methods of harassment, sophisticated attacks, and technological advances in hate speech” [20].

AI can fuel and vastly multiply online smear campaigns and accelerate the dissemination of misconceptions and misinformation about the LGBTIQ+ communities. Tech companies are slow to rise to this challenge and a lot of the harm being done against individuals and the LGBTIQ+ communities remains largely unaddressed and unaccounted for [21].

1.1.2 Threats for physical health and safety

AI is also applied in tools that generate medical and psychological advice but has been found to generate incorrect medical advice [22]. In a world where gender non-conforming bodies and minds remain massively understudied and most medical data is mostly developed and tested on male, cisgender, straight, white bodies [23], automating medical advice is likely to misinform members of the LGBTIQ+ communities [24-25]. Such developments are even more harmful for LGBTIQ+ persons, as they rely more strongly on online spaces for information on physical and mental health [26].

Moreover, there appears to be little to no legislation as to who regulates such programs and where they can or cannot be used. Although the latest provisional versions of the EU AI Act mention the need to regulate the use of AI in “essential private services and public services” [27], a gap would remain as to who may develop and control such platforms. This would create a risk of such essential information channels being hijacked by FIMI Actors.

1.1.3 Fuelling bias and hate

The facilitated production of disinformation against LGBTIQ+ communities does not only affect LGBTIQ+ communities directly but may also increase stigma of an already hostile population through confirmation bias: when someone would like a certain idea or concept to be true, they end up believing it to be true [28]. As stated in the 2023 EEAS report on FIMI targeting LGBTIQ+ communities, “disinformation and other types of manipulative content do not necessarily lead to bias against certain social groups but rather resonate with audiences in which certain biases are already deeply ingrained” [29].

FIMI actors also target countries in which an anti-Western sentiment and discourse is prevalent, feeding into the narrative that the fundamental rights of LGBTIQ+ people rights are “propaganda from the West”. Examples include a fake flyer with two men kissing that circulated amongst African social media accounts [30], or a fake tweet by Biden calling out the Ugandan Parliament on introducing the LGBTIQ+ bill proposed in 2023 [31]. Such content is instrumentalised by Russia in Africa to gain more public and military support against Europe [32].

Such developments tie into a global increase in anti-LGBT policies and laws. A record number of anti-LGBTIQ+ policies have been introduced and implemented in the last year—and they’re having a negative impact on LGBTIQ+ young people’s mental health [33]. Nearly 2 in 3 LGBTIQ+ young people said that hearing about potential state or local laws banning people from discussing LGBTIQ+ people at school made their mental health a lot worse [34].

Moreover, the advent of AI can level up prosecution tactics [35] of homophobic governments to monitor and punish LGBTIQ+ individuals with unprecedented speed and sophistication. The Russian government has already launched [36] an AI-driven system aimed at identifying “illegal” content online to enforce the “gay propaganda” law [37].

Although the publication and creation of hateful content against the LGBTIQ+ communities are constant, defamation or smear campaigns most often take place during important solidarity events for LGBTIQ+ communities, such as Pride events and marches. Other public events such as protests, assassinations or elections have also been found to coincide with an increase of online hate [38] and in several instances AI has facilitated this.

Examples targeting the LGBTIQ+ communities include the false belief that the Arc de Triomphe was projected with a rainbow for Pride [39], misrepresentations of the LGBTIQ+ communities [40] or generally increasing the amount of hateful content during Pride. Identifying such patterns of high risk can be used to increase efforts against online hate and disinformation in a more targeted manner.

1.2 Freedom of expression

1.2.1 Unsafe spaces

Freedom of speech is not equally and actively protected when there is a lack of concern and care regarding anti-LGBTIQ+ hate speech online. In 2023, the international LGBTIQ+ organisation Trevor Project, working on suicide prevention amongst the LGBTIQ+ communities, left X (formerly known as Twitter) because of the suppression of content moderation standards. It stated that it had therefore become too difficult to create a safe space for LGBTIQ+ youth on this platform [41]. This is an example of how hateful content can directly impact the effective freedom of expression for everyone and especially online voices advocating for the rights of LGBTIQ+ communities.

This example is extremely worrying when taking into consideration the possible multiplication of hateful content through generative AI, as authentic countervoices may be not only smothered, but also decide to retreat from online spaces due to the difficulty to face this violence, affecting these countervoices' freedom of expression by increasing self-censorship. There are other examples of individual human rights defenders who avoid announcing online that they work on LGBTIQ+ issues, out of fear for retaliation against their work [42].

1.2.2 Insufficient and biased moderation

Automated content moderation presents a risk for freedom of expression, especially when the algorithms implementing content moderation policies are trained on biased datasets. It is even more problematic when those algorithms are implementing anti-LGBTIQ+ policies. The owner of X (formerly known as Twitter), Elon Musk, removed rules in X's speech guidelines that banned deliberate "misgendering" and "deadnaming" of transgender users [43]. He also categorised the terms "cis" and "cisgender" as slurs [44]. This is especially dangerous because such classifications can easily be automated, which immediately and systematically restricts the freedom of speech of LGBTIQ+ persons and their allies online [45].

Moreover, a group of researchers and YouTubers found that algorithms behind automated content moderation appear biased against LGBTIQ+ content: "Our testing result clearly showcase[s] that perfectly acceptable titles that are otherwise perfectly fine for monetization are demonetized only when 'gay' and 'lesbian' are added to the title" [46]. Other analyses also show biases against the word 'gay', 'lesbian', 'bisexual', 'sexual' 'drag queens' or 'same sex marriage' [47].

Although X has been found to be the worst social media platform for countering anti-LGBTIQ+ content and general disinformation, all major social media platforms lack effective mechanisms to protect LGBTIQ+ communities online. In their 2023 Social Media Safety Index, GLAAD recommended specifically to refrain from AI-led content moderation, pointing out its specific bias against marginalised groups [48-49].

Indeed, automated content moderation appears to work well for specific uses like image copyright or child pornography but is less successful in detecting hateful text-based content [50]. Meta uses a mix of human and AI-led content moderation and applies human-led or AI-led content moderation depending on how content is flagged by users [51].

However, the transparency report does not include clear data on what proportion of content moderation concerning LGBTIQ+ content is done by humans and AI [52]. In 2020, Meta and Twitter did introduce “speedbumps” for certain harmful content going viral, which have been used effectively to slow the spread of misinformation, including anti-LGBTQ hateful content [53].

Gaps in content moderation policies even more strongly impact non-English languages. A recent report by Localisation Lab found that content moderation reports are extremely poorly translated, and that there were several barriers for non-English speakers to effective interpretation and contextualisation of content [54]. A DSA transparency report [55] on X showed that the platform hardly hired any content management personnel whose primary language was not English (see image). This means that the already flawed content moderation policies will fail even more strongly in eliminating hateful content in non-English languages.

Primary Language	People
Arabic	12
Bulgarian	2
Croatian	1
Dutch	1
English	2,294
French	52
German	81
Hebrew	2
Italian	2
Latvian	1
Polish	1
Portuguese	41
Spanish	20

Until now, automated content generation and moderation tools continue to fail to implement inclusive and non-sexist language and continue to reproduce gender stereotypes [56]. To effectively counter LGBTIQphobia online and on a global scale, the active deconstruction of stereotypes and biases, and the renewal of language standards is needed. For this, the active inclusion of LGBTIQ+ persons and experts in developing, implementing, and evaluating platforms and content moderation policies is needed [57].

Reports identifying practices for gender-inclusive content moderation quickly flag the risk of bias in automated approval and disapproval of content [58]. At the same time, the traumatising effect of violent and graphic content on human content moderators should be considered [59]. Lessened workload or the help of AI-powered content moderation tools can help reduce mental health consequences on content moderators.

1.3 Non-discrimination

1.3.1 Discriminatory AI

As we pointed out earlier, AI is not able to detect bias without human intervention and easily reproduces and reinforces biases that human rights movements have been working to quash. The use of AI in data collection and research, for example, risks amplifying existing biases and risks enforcing scientific conclusions that exclude gender non-conforming persons or spread misinformation [60].

AI often fails in taking into consideration contextual factors and nuance and can therefore easily reproduce harmful stereotypes about LGBTIQ+ people when fed with biased data [61]. Biases and stereotypes are also found in AI-led predictive policing programs, which have important bias against marginalised populations in governmental and legislative institutions [62-63]. Such use of AI can easily feed into FIMI strategies against LGBTIQ+ communities.

Generative AI also frequently reproduces sexist or binary language [64]. New developments like inclusive languages or non-binary pronouns are hardly considered, even upon request [65]. Even the personalities of bots and voice assistants reproduce gender bias [66]. Another example is when a creator used Midjourney to generate 100 gay couples. They noticed that “almost every single couple is white, young, and skinny” [67]. An analysis of more than 5,000 images created with Stable Diffusion found that it “takes racial and gender disparities to extremes — worse than those found in the real world” [68].

1.3.2 AI undermining LGBTIQ+ identities

While the pseudoscientific allure of gender, sexual orientation and emotional recognition technologies proliferate, stereotypes and prejudices can be further entrenched into datasets and LGBTIQ+ persons have less autonomy over their “algorithmic identity”. As explained by Access Now: “Our private spaces are also being opened to systems that purport to detect gender. Spotify was recently granted a speech-recognition patent [69] for a system that claims to detect [70], among other things, your “emotional state, gender, age, or accent” to recommend music. On April 2, we sent a letter to Spotify [71] calling on the company to abandon the technology.”

The Spotify patent reflects a broader trend of subtle but impactful uses of AI, like gender categorization systems, shaping a new “algorithmic identity.” This

integration reinforces flawed technologies, perpetuating discrimination against trans and non-binary individuals by erasing their identities in digital spaces and normalising this erasure, as argued by Daniel Leufer of Access Now [72].

1.3.3 AI, LGBTIQ+ discrimination and FIMI

Algorithmic discrimination can significantly impact LGBTIQ+ individuals within the framework of FIMI, but this requires more research. When foreign actors create anti-LGBTIQ+ campaigns to sow discord, spread false narratives, or incite prejudice against LGBTIQ+ individuals, they may be feeding additional generative AI models that will continue to reproduce those results.

An important risk presents itself when FIMI actors contribute to data feeding AI-led tools that are used for real-life situations, selection procedures or implementation of public services. AI is used in recruitment tools [73], medical diagnostic tools [74] or border crossings [75]. Such tools pose a threat to marginalised populations, especially when they are fed with discriminatory data or influenced by hostile (or FIMI) actors.

Addressing LGBTIQ+ algorithmic discrimination within the context of FIMI requires developing more robust algorithms that have been audited by experts on LGBTIQ+-related topics to mitigate biases, enhancing digital literacy to help individuals critically evaluate information, and fostering collaboration between governments, tech companies, civil society, and LGBTIQ+ organisations to identify and counter anti-LGBTIQ+ mis- and disinformation effectively.

1.4 Privacy

Online privacy is particularly important for LGBTIQ+ individuals. Considering the widespread stigma against LGBTIQ+ communities and individuals who are openly part of those communities, it is particularly important for LGBTIQ+ persons to be able to control what part of their life and identity is public and what is private [76]. Nevertheless, the right to privacy of LGBTIQ+ people is threatened by several AI developments that need to be addressed.

1.4.1 Data breaches

Data breaches can disproportionately impact the LGBTIQ+ communities [77], since the consequences of revealing the gender identity or sexual orientation of an LGBTIQ+ person can, in some contexts, create a direct risk for their life and safety due to persisting stigmatisation of LGBTIQ+ persons worldwide.

Data breaches happen as tech companies are collecting a massive amount of data of their users, including LGBTIQ+ users (including search histories [78] or social activity online) which can cause serious problems if this data is leaked or harvested to target LGBTIQ+ communities.

In 2017, a large amount of visual data on transgender individuals was collected without their consent to train facial recognition software [79]. In 2018, the dating app Grindr experienced a data leak, after sharing users' data on their HIV status with third parties [80-81]. The company was systematically collecting and displaying this information to contribute to a safe space by destigmatising HIV status, but the consequences of this data not being sufficiently protected completely unravelled that progress, putting users at risk.

The Wall Street Journal has since reported that Grindr sold over 13 million users' data to the highest bidder between 2017 and 2020 [82]. Even though the company has since changed its policy, some of that data may still be available for sale. This data (which includes geolocation data) can be weaponized by actors with nefarious intent (e.g. FIMI actors). Privacy issues are also raised concerning social media platforms like TikTok, especially when it comes to data collection for targeted advertising and tailored suggestions in feeds [83].

1.4.2 Offline surveillance

Pseudoscientific automated gender recognition (AGR) tools, including tools that are marketed as being able to identify a person's sexual orientation identity or gender identity [84], pose great risks to offline civic space and the ability of the LGBTIQ+ communities to safely participate in public life. Widely used facial recognition technologies include the classification of the gender identity of trans and non-gender-conforming individuals [85].

Such classifications are impossible to make directly from physical traits (facial traits, hair, body shape) or elements of gender expression (dress, make up...) or by gathering data on someone, as sexual orientation and the gender identity are inner perceptions that (1) only exist on a spectrum and are not binary or classifiable, (2) can only really be known to the persons themselves and (3) may change over time.

The development of such AI-powered tools can be harmful for LGBTIQ+ persons in many ways [86]. AGR programs put at risk gender non-conforming individuals, as they are often misgendered or flagged as "anomalies". When such softwares are used in recruitment processes or border controls, persons may be forced to explain their personal history, which puts them at unnecessary risk [87].

Facial recognition cameras with AGR impact several rights, including peoples' ability to move freely, access public services [88] and engage in public protests without fear of threats or attack [89].

Moreover, the data produced by such programs may be used to register gender non-conforming individuals, creating data that could be used against them by hostile (FIMI) actors. Finally, such tools may be used to falsely associate public figures with the LGBTIQ+ communities, creating a risk of stigma against these figures. This also risks framing non-LGBTIQ+ persons as representing the community and providing them with 'legitimacy' to speak in the name of LGBTIQ+ people. This creates a risk for accurate representation of LGBTIQ+ communities in the public and political space.

Another example is surveillance programs used in schools, monitoring students' activity online. Surveillance software communicates data on online search histories to teachers, alert them of "riskful" activity and block students from accessing certain websites [90]. Such softwares were found to frequently "flag harmless activity as a 'threat,'" and frequently prevent queer or queer curious students accessing information "by flagging words related to sexual orientation" [91]. It was established that "students from minority or marginalized communities, including students of colour and LGBTIQ+ students, are far more likely to be flagged" [92]. Such programs may forcibly "out" students and limit their exploration of their personal identity online. Moreover, such programs treat very sensitive data that, when falling into the wrong hands, could easily be used against them.

Overall, AI-led technologies erode the privacy, informed consent, and agency of LGBTIQ+ persons to disclose their gender identity and sexual orientation identity on their own terms [93]. For both domestic and foreign actors, such data is extremely valuable, especially to those who are looking to influence public policies or even electoral processes [94]. This brings into question the necessity and proportionality of these tools [95].

1.5 Democratic processes and international security

1.5.1 Right to free and fair elections

The intersection of misinformation, AI and LGBTIQ+ topics also has an important impact on political campaigns [96]. With AI, it becomes easier to create more credible misrepresentative content on politicians aimed at tarnishing their reputation. The Freedom House report sheds light on these risks, emphasising the need for vigilant measures in safeguarding democratic processes [97]. The Center for Democracy & Technology also raised concerns about the risks posed by fraudulent misrepresentations in political campaigns facilitated by AI [98].

There are already examples of deceptive impersonation videos [99], as well as examples of spreading gay rumours about political candidates [100]. There are several examples of falsely stating a political candidates' support for LGBTIQ+ communities, such as Mukwege in the DRC [101] or Nigeria's opposition presidential candidate, Atiku Abubakar in 2018 [102]. AI will make it much easier to create false associations of politicians with the LGBTIQ+ communities during electoral campaigns by accompanying mis/disinformation with hyper-realistic images and deep fake videos.

The false linking of public figures to the LGBTIQ+ communities also takes place outside an electoral context but continues to influence their political reputation on the international stage. Examples include Pope Francis wearing the Rainbow Flag [103], a fake image of Zelensky at a Pride event in 1999 [104], or a video of Joe Biden pronouncing violent discourse against the LGBTIQ+ communities [105].

In the US, it appears that tech companies and social media platforms are going to rely increasingly on AI to monitor content around elections because of lay-offs in content moderation personnel [106]. Considering the important role of LGBTIQ+ voters in these elections (as mentioned earlier) combined with the biased nature of AI, there is a grave risk that LGBTIQ+ communities' content will be more likely to be censored, impacting online political discourse.

Moreover, data can be used to target specific audiences to influence voter opinions or voting behaviour. For example, LGBTIQ+ voters may receive text messages or messages on social media saying that it is not possible or safe for them to vote in certain places. This already happens offline: in the Democratic Republic of the Congo (DRC), LGBTIQ+ persons are already being refused their electoral card or face physical violence when getting them [107]. Targeted harassment of LGBTIQ+ persons during election time is a real and tangible risk to democratic values and the right to fair elections.

When there is a surplus of available data online paired with powerful digital investigative tools, members of the LGBTIQ+ communities can be more easily targeted with false or misleading information about how to vote or when elections take place, making it easier to manipulate their understanding of whether or not they have actually voted [108]. In the US, data shows that the LGBTIQ+ communities are an increasingly powerful voting group [109]. The targeting of this group by national and foreign actors could significantly diminish this power and threaten the effective participation of LGBTIQ+ people in elections. In Europe, this same risk exists.

1.5.2 International peace and security

FIMI can feed into igniting conflict, but also exacerbate harms during conflict, particularly to historically marginalised groups. Misinformation has long been used in genocidal conflicts [110] but the current access and dependency of civilians on online information increases the reach and influence of online misinformation in conflict [111]. Online hate and misinformation can be used to limit certain group's access to health services, for example, or increase the stigmatisation of specific groups in an already precarious situation [112]. The online space, social media and information warfare have become an integral part of conflict dynamics [113-114] and an integral part of warfare strategies [115].

Artificial intelligence adds complications to already difficult situations for LGBTIQ+ persons by facilitating the spread of misinformation and facilitating the targeting of specific minority groups. During conflict, members of the LGBTIQ+ communities already often face additional risks as compared to persons outside these communities [116]. LGBTIQ+ persons receive less close community support and face additional difficulties in accessing their medical needs [117].

To meet those needs, LGBTIQ+ persons generally rely more strongly on online spaces for coordinating such support, for example to organise meetings or to facilitate access to gender-related care [118]. Therefore, the intervention of FIMI and AI not only risks increasing exposure to hate online but could also facilitate the intervention of actors with online-facilitated access of LGBTIQ+ persons to community support and health services. Misinformation (e.g. impersonification) could be used, for example, to direct those communities towards specific spaces and to organise targeted attacks [119].

In conflict situations, certain organisations have already noted an increased trust and use of certain social media platforms (e.g. Telegram) for essential health services and emergency support [120]. FIMI actors make use of this development to spread disinformation about evacuation information or other essential services, directly putting at risk specific populations in precarious situations. Because LGBTIQ+ communities more heavily depend on online spaces for essential services (e.g. for information on queer-related health services or for finding community within an anti-LGBTIQ+ context), they become an easy target of such attacks. AI makes such campaigns easier and faster to implement and therefore increases such risks.

ISIS, for example, has been reported to use social media to track down members of the gay community for killings [121]. AI and FIMI risk intensifying and facilitating such practices of targeting the LGBTIQ+ communities in conflict settings. Online, LGBTIQ+ topics and AI are being used to influence the global public opinion on the conflict. For example, many Russia-led social media outlets spread misinformation linking the Ukrainian army to the LGBTIQ+ communities to damage their reputation [122]. In September 2022, a photo of Zelensky at a Pride event, which turned out to be photoshopped [123], appeared on Twitter. Other posts of him at other Pride events also went around on Facebook and Instagram.

Israel is currently developing a "pinkwashing" narrative, claiming that crimes carried out by their army is to save LGBTQI+ persons from an oppressive and homophobic regime in Palestine [124]. Recently a video was put online of people being thrown off a building, which was claimed to be depicting Palestinian LGBTIQ+ persons being thrown off buildings in Palestine while it was depicting a Syrian Islamic State jihadist group [125]. AI poses an additional threat as it facilitates the production of such images and content, fuelling the use of anti-LGBTIQ+ narrative in influencing the public opinion on conflict dynamics.

The development of AI in misinformation also poses a risk in international relations and global power dynamics: Russia is already using disinformation to influence public opinion and political alliances in Africa, for example [126]. Furthermore, AI creates a foreseeable risk of automatised and specific targeting of LGBTIQ+ people through spyware software or through automated weaponry [127].

2. WAYS FORWARD

2.1 Developments and opportunities

The EU has developed several legislations that offer opportunities for regulation of the use of AI. The Digital Services Act (DSA), introduced in 2022, attempts to regulate online platforms. It focuses on different elements like the prohibition of “illegal content” from big online platforms. However, a report by EDPS notes that more detail was needed regarding content moderation, implementation of data protection measures, handling reports and complaints and other elements [128].

Regarding FIMI, the DSA leaves an important gap because it focuses solely on illegal content, not illegitimate or harmful content [129]. The DSA does include the Code of Practice on Disinformation, although a recent evaluation has shown that “most of the platforms are not implementing the measures they committed to” [130].

Another opportunity for advocacy is the EU AI Act, which is the first regulation on AI soon to be adopted [131]. Although the final text is yet to be approved, a provisional political agreement was reached in December 2023 between the Council of the EU and the European Parliament [132].

Following this provisional agreement, many uses of AI are expected to be prohibited in the EU, like AGR biometric categorisation systems (including those who target sexual orientation), the untargeted scraping of facial images from the internet or CCTV footage to create facial recognition databases, programs using emotion recognition in the workplace and educational institutions, programs providing social scoring, manipulative AI or AI exploiting the vulnerabilities of people.

Although such provisions sound hopeful if they are to be adopted, it remains to be seen what is exactly understood under these different elements and whether the risks listed in this report are adequately addressed by this legislation [133].

Many big tech companies have been publishing their own voluntary “responsible AI” principles – Google [134], Microsoft [135], Meta [136], etc – but the EU needs to make sure that these are held accountable by binding obligations, and that they are operating in accordance with international human rights law and industry best practices. Many tech companies are outwardly committing to these principles while internally cutting down their human rights [137] and responsible AI teams [138], bringing into question the legitimacy of their “human centred” AI claims.

There are countless examples of how the failure of tech companies to effectively design, develop, distribute, and implement safe products (that have evolved through a human rights due diligence process) have disproportionately impacted historically marginalised groups, and this report merely scratches the surface. We need government action—for example, by making sure the EU AI Act is strong and with teeth—to address the nuanced challenges of the digital age.

More research is needed to fully understand the extent to which AI-augmented FIMI is a threat to European democracies, as well as LGBTIQ+ communities. Identifying the sources of FIMI is already difficult, and AI adds additional layers of complications with accountability [139]. What is clear, is that these salient human rights risks cannot be ignored, considering the many examples of how AI is already contributing to harms.

2.2 Conclusion and recommendations

In an era where the rapid evolution of Artificial Intelligence (AI) intersects with the delicate fabric of human rights, particularly those of LGBTIQ+ communities, the need for thoughtful and proactive measures has never been more critical. The emergence of AI as a powerful tool in shaping public opinion and discourse is accompanied by many challenges, especially in the context of disinformation campaigns. These campaigns, often orchestrated by FIMI actors, pose a significant threat to the safety, dignity, and equality of LGBTIQ+ individuals.

Our recommendations are crafted with a deep understanding of the nuanced ways in which AI can be wielded as a tool for both harm and good. They are guided by a commitment to uphold human rights, promote inclusivity, and counteract the pernicious effects of disinformation. As we chart a path forward, it is our collective responsibility to ensure that the advancement of AI technology aligns with the principles of equality, respect, and dignity for all, particularly for those within the LGBTIQ+ communities who are often at the frontline of digital vulnerabilities.

Following the analysis presented in this report on the potential harmful risks presented by AI-systems regarding the protection of the fundamental rights of LGBTIQ+ people, we demand that the EU institutions and national governments:

- Develop stricter and explicit regulations for social media platforms to actively create a safe environment for LGBTIQ+ persons and organisations. Make sure that experts on LGBTIQ+ topics are involved to carry out sensitivity checks in different languages and provide expertise to online platforms on LGBTIQ+ specific challenges.
- Develop stricter regulations for data protection by online platforms, with a special focus on LGBTIQ+ related data.
- Require social media companies to cooperate with civil society organisations to moderate anti-LGBTIQ+ disinformation, and to support tracking of the use of AI tools for anti-LGBTIQ+ disinformation.
- Require transparency concerning the social media platforms' content moderation systems and recommendation algorithms, clarifying what content moderation is done by humans and what content moderation is done by AI. Increase human content moderation specifically for anti-LGBTIQ+ content with the support of the expertise of human content moderators on LGBTIQ+ related topics.
- Ensure that tech companies have effective grievance mechanisms to facilitate access to remedy when LGBTIQ+ persons experience harm by AI-powered technologies.
- Define clear sanctions to hold social media companies and other service providers fully accountable for their failure to effectively and efficiently moderate harmful content on their platforms.
- Ensure that facial recognition tools claiming to recognise a person's gender identity or sexual orientation are not able to be designed, developed, or deployed in the EU or exported to other countries [140]
- Ensure that the needs of the LGBTIQ+ communities are explicitly reflected in the fundamental rights impact assessments that are expected to be required within the EU AI Act and the CSDDD, especially during the design, development, and sale of technologies. Make sure that LGBTIQ+ civil society organisations are involved in these processes.

- Require online platforms to collect more targeted data on where and when harmful content targeting LGBTIQ+ communities is produced, including language and country specific data. Provide civil society and monitoring mechanisms with access to this data, to collaborate against anti-LGBTIQ+ content more effectively.

Beyond these policy recommendations, there is a critical need to make both public and private funding available to:

- Support additional research to better understand how FIMI and the targeting of LGBTIQ+ people is impacting both online and offline civic space, which is necessary to preserve a healthy, democratic, and inclusive information ecosystem.
- Support additional research on AI bias against LGBTIQ+ people and communities.
- Increase the monitoring of online AI-facilitated FIMI against the LGBTIQ+ communities. This includes but is not restricted to monitoring content during Pride events, elections, political protests and hate crimes targeting LGBTIQ+ people. An example of good practice can be the implementation of ‘speedbumps’ regarding certain topics and events, slowing down the spreading of specifically harmful content [141]. Such monitoring requires a coordinated approach at the European level involving local civil society, national governments, and European institutions.

Regarding these last recommendations, it is crucial to make sure that the funding provides for the full involvement of LGBTIQ+ civil society organisations in conducting such monitoring and research activities.

References

- [1] <https://edmo.eu/2023/05/30/rights-in-the-time-of-conspiracies-and-fake-news-disinformation-against-lgbtq-in-the-eu/>
- [2] <https://www.eeas.europa.eu/sites/default/files/documents/2023/EEAS-LGBTQ-Report-03-Digital%201.pdf>
- [3] <https://www.egmontinstitute.be/countering-fimi-a-critical-imperative-for-mission-safety/>
- [4] <https://orbofi.medium.com/decentralizing-content-creation-how-ai-is-democratizing-the-creative-process-199938873b29>
- [5] Broussard, M. (2023). More than a glitch: Confronting Race, Gender, and Ability Bias in Tech. The MIT Press
- [6] <https://www.ibm.com/topics/ai-hallucinations>
- [7] <https://counterhate.com/research/misinformation-on-bard-google-ai-chat/>
- [8] Broussard, M. (2023). More than a glitch: Confronting Race, Gender, and Ability Bias in Tech. The MIT Press
- [9] <https://www.pewresearch.org/internet/2019/09/05/the-challenges-of-using-machine-learning-to-identify-gender-in-images/>
- [10] <https://www.codastory.com/authoritarian-tech/ai-sexuality-recognition-lgbtq/>;
<https://www.nytimes.com/2017/10/09/science/stanford-sexual-orientation-study.html>
- [11] Broussard, M. (2023). More than a glitch: Confronting Race, Gender, and Ability Bias in Tech. The MIT Press
- [12] An exception is the 2023 EEAS report https://www.eeas.europa.eu/eeas/fimi-targeting-lgbtq-people_en
- [13] <https://www.ilga-europe.org/press-release/deadliest-rise-anti-lgbti-violence-decade/>
- [14] <https://www.npr.org/2021/05/10/995328226/social-media-hate-speech-harassment-significant-problem-for-lgbtq-users-report>
- [15] https://s3.us-west-2.amazonaws.com/media.glaad.org/wp-content/uploads/2021/05/20104937/GLAAD-SOCIAL-MEDIA-SAFETY-INDEX_Q.pdf
- [16] <https://www.washingtonpost.com/news/the-switch/wp/2014/03/05/the-internet-is-both-a-blessing-and-a-curse-for-ugandas-lgbti-activists/>
- [17] <https://www.hrw.org/report/2023/02/21/all-terror-because-photo/digital-targeting-and-its-offline-consequences-lgbt>
- [18] ILGA also highlighted the specific impact of this violence on LGBTIQ+ youth: “Over half of LGBTI students in Denmark had suicidal thoughts or self-harmed, while 82% of LGBTI students in Northern Ireland thought of suicide. 80% of LGBT students feel unsafe in school in Ukraine and 40% missed school only this past month because of this”. A 2023 survey by the Trevor Project found that “discrimination and online hate contributes to higher rates of suicide risk reported by LGBTQ young people”. Other reports confirm that young LGBTIQ+ people are more than twice as likely to experience hate speech online compared with those who identify as heterosexual.
- [19] <https://orca.cardiff.ac.uk/id/eprint/127085/1/Hate%20Behind%20the%20Screens.pdf> (p.18); <https://galop.org.uk/resource/online-hate-crime-report-2017/>
- [20] <https://unesdoc.unesco.org/ark:/48223/pf0000387483>
- [21] “Ads submitted to Facebook, TikTok and YouTube containing extreme violent hate against the LGBTQ+ community in Ireland were almost all approved for publication by the platforms, according to a new test carried out by Global Witness.” (April 2023) <https://www.globalwitness.org/en/press-releases/extreme-and-violent-anti-lgbtq-hate-approved-publication-leading-social-media-platforms/>
- [22] <https://pubmed.ncbi.nlm.nih.gov/37881016/>
- [23] <https://www.himss.org/resources/uncovering-and-removing-data-bias-healthcare>
- [24] <https://www.forbes.com/sites/jamiewareham/2021/03/21/why-artificial-intelligence-will-always-fail-lgbtq-people/>
- [25] Broussard, M. (2023). More than a glitch: Confronting Race, Gender, and Ability Bias in Tech. The MIT Press
- [26] https://link.springer.com/chapter/10.1007/978-3-319-69638-6_4
- [27] <https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>
- [28] <https://www.britannica.com/science/confirmation-bias>
- [29] <https://www.eeas.europa.eu/sites/default/files/documents/2023/EEAS-LGBTQ-Report-03-Digital%201.pdf>
- [30] https://www.eeas.europa.eu/eeas/fimi-targeting-lgbtq-people_en
- [31] <https://www.reuters.com/article/idUSL1N3670R0/>
- [32] <https://edmo.eu/2023/11/29/what-russian-disinformation-wants-africa-to-think-about-the-war-in-ukraine-and-more/>
- [33] <https://www.thetrevorproject.org/survey-2023/#anti-lgbtq-policies>
- [34] <https://www.thetrevorproject.org/survey-2023/#anti-lgbtq-policies>
- [35] <https://www.article19.org/wp-content/uploads/2022/03/Digital-Crime-Scenes-Report-3.pdf>
- [36] <https://www.reuters.com/world/europe/russian-system-scan-internet-undesired-content-dissent-2023-02-13/>
- [37] <https://www.reuters.com/world/europe/russian-system-scan-internet-undesired-content-dissent-2023-02-13/>
- [38] For example, during protests of Black Lives Matter after the assassination of George Floyd, “hate speech targeting gender identity and sexual orientation – a topic with little intuitive connection to the murder and protests – increased by 75%.” <https://www.theguardian.com/society/2023/jan/25/real-world-events-trigger-online-hate-toward-unrelated-groups-study-finds>
- [39] <https://www.reuters.com/article/factcheck-pride-idUSL1N37U2E7/>
- [40] <https://www.mediamatters.org/facebook/instagram-allowing-accounts-spew-hate-lgbtq-people-while-also-claiming-support-community>
- [41] <https://mashable.com/article/trevor-project-lgbtq-organizations-leaving-twitter-x>
- [42] <https://www.protectioninternational.org/researchpublications/the-right-to-defend-human-rights-not-for-all-drc/>
- [43] <https://globalnews.ca/news/9783640/elon-musk-cisgender-twitter/>
- [44] <https://www.forbes.com/sites/kimelsesser/2023/07/02/elon-musk-deems-cis-a-twitter-slurheres-why-its-is-so-polarizing/>
- [45] Many accounts have been identified and reported but continue to be accepted on different social media platforms#. X is often mentioned to be fuelling hateful content, continuing to “host 86% of 300 posts reported for extreme hate speech”. <https://counterhate.com/research/twitter-x-continues-to-host-posts-reported-for-extreme-hate-speech/>
- [46] <https://www.vox.com/culture/2019/10/10/20893258/youtube-lgbtq-censorship-demonetization-nerd-city-algorithm-report>
- [47] https://info.cheg.ai/hubs/Research/Brand_Safety_Blocklist_Report
- [48] <https://assets.glaad.org/m/7adb1180448da194/original/Social-Media-Safety-Index-2023.pdf>
- [49] <https://assets.glaad.org/m/7adb1180448da194/original/Social-Media-Safety-Index-2023.pdf>
- [50] <https://teseo.unitn.it/cll/article/view/2365>
- [51] https://www.lemonde.fr/en/pixels/article/2023/11/14/content-moderation-key-facts-to-learn-from-facebook-instagram-x-and-tiktok-transparency-reports_6252988_13.html#
- [52] <https://teseo.unitn.it/cll/article/view/2365>
- [53] <https://slate.com/technology/2020/08/facebook-twitter-youtube-misinformation-virality-speed-bump.html>
- [54] <https://www.localizationlab.org/tech-policy-review>
- [55] https://twitter.com/romanadamczyk5/status/1720451369004401087?s=46&t=PAq7NQCvt_qDKVzd7VcAlw

- [56] <https://www.cigionline.org/articles/generative-ai-tools-are-perpetuating-harmful-gender-stereotypes/>
- [57] <https://www.forbes.com/sites/forbestechcouncil/2023/09/01/why-businesses-need-to-embrace-inclusive-content-moderation-strategies/>
- [58] https://assets-global.website-files.com/641dc6058ca7b7b65422b5bd/64593fc0721e2b44b1d1d9b5_CX-White%2BPaper-2021.pdf
- [59] <https://time.com/5405343/facebook-lawsuit-mental-trauma-content-ptsd/>
- [60] Broussard, M. (2023). More than a glitch: Confronting Race, Gender, and Ability Bias in Tech. The MIT Press
- [61] <https://aquariusai.ca/blog/why-artificial-intelligence-fails-uncovering-the-limitations-of-ai-technology>
- [62] Broussard, M. (2023). More than a glitch: Confronting Race, Gender, and Ability Bias in Tech. The MIT Press
- [63] <https://onlinelibrary.wiley.com/doi/full/10.1002/ajcp.12671>
- [64] <https://www.scientificamerican.com/article/chatgpt-replicates-gender-bias-in-recommendation-letters/>
- [65] <https://www.machinetranslation.com/blog/creating-an-inclusive-ai-future-the-importance-of-non-binary-representation>
- [66] <https://www.brookings.edu/articles/how-ai-bots-and-voice-assistants-reinforce-gender-bias/>
- [67] <https://gayety.co/this-is-what-happens-when-ai-generates-100-gay-couples>
- [68] <https://www.bloomberg.com/graphics/2023-generative-ai-bias/>
- [69] <https://patents.justia.com/patent/10891948>
- [70] <https://www.bbc.com/news/entertainment-arts-55839655>
- [71] <https://www.accessnow.org/spotify-tech-emotion-manipulation/>
- [72] <https://www.accessnow.org/how-ai-systems-undermine-lgbtq-identity/>
- [73] <https://www.pewresearch.org/internet/2023/04/20/ai-in-hiring-and-evaluating-workers-what-americans-think/>
- [74] <https://arxiv.org/abs/2401.05654>
- [75] <https://www.theguardian.com/us-news/2023/apr/03/us-mexico-border-surveillance-towers-customs-border-protection>
- [76] <https://staysafeonline.org/online-safety-privacy-basics/data-privacy-crucial-lgbt-community/>
- [77] <https://staysafeonline.org/online-safety-privacy-basics/data-privacy-crucial-lgbt-community/>
- [78] <https://www.business-humanrights.org/en/latest-news/openai-takes-chatgpt-offline-due-to-data-breach/>
- [79] <https://www.theverge.com/2017/8/22/16180080/transgender-youtubers-ai-facial-recognition-dataset>
- [80] <https://www.theguardian.com/commentisfree/2018/apr/04/grindr-gay-men-hiv-status-leak-app>
- [81] <https://www.theguardian.com/technology/2018/apr/03/grindr-shared-information-about-users-hiv-status-with-third-parties>
- [82] <https://www.wsj.com/articles/grindr-user-data-has-been-for-sale-for-years-11651492800>
- [83] https://www.amnesty.org.pl/wp-content/uploads/2023/11/TikTok_I_feel_exposed.pdf
- [84] <https://www.theguardian.com/technology/2017/sep/07/new-artificial-intelligence-can-tell-whether-youre-gay-or-straight-from-a-photograph>
- [85] <https://law.unimelb.edu.au/news/caide/machines.-artificial-intelligence-and-the-rising-global-transphobia>
- [86] For context, Access Now explains: “Numerous studies and audits have shown that facial recognition based AGR technology is not accurate for many people. In their groundbreaking study, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, Joy Buolamwini and Timnit Gebru found that the AGR systems deployed by prominent companies had higher error rates for women than men, and the failure rate was even higher for dark-skinned women. Unfortunately, making facial recognition more accurate would not diminish its discriminatory impact in many other contexts — and as we’ll show, it’s highly unlikely any adjustments would make it any less harmful for trans and non-binary people.”
- [87] <https://www.advocate.com/transgender/2016/3/04/new-tsa-rule-bad-news-trans-people-says-task-force>
- [88] See the Privacy International piece: “Research from Coding Rights and Privacy International has also mapped the use of AGR to authenticate IDs to access public services in Brazil, including a survey where they found that “90.5% of trans people responded that they believe facial recognition can operate from a transphobic perspective; 95.2% had the impression that this technology can leave them vulnerable to situations of embarrassment and contribute to the stigmatization of trans people.”
- [89] <https://www.popularmechanics.com/technology/security/a32851975/police-surveillance-tools-protest-guide/>
- [90] <https://youthtoday.org/2022/11/how-the-rise-of-school-surveillance-software-affects-lgbtq-students/>
- [91] <https://youthtoday.org/2022/11/how-the-rise-of-school-surveillance-software-affects-lgbtq-students/>
- [92] <https://youthtoday.org/2022/11/how-the-rise-of-school-surveillance-software-affects-lgbtq-students/>
- [93] See the case of the “facial recognition trans database”
- <https://www.vice.com/en/article/93aj3z/facial-recognition-researcher-left-a-trans-database-exposed-for-years-after-using-images-without-permission>
- [94] <https://ourdataourselves.tacticaltech.org/posts/inside-the-influence-industry/>
- [95] See the example from Access Now: “Consider an advertising billboard that “detects” your gender and switches from advertising power tools for “men” to summer dresses for “women.” Not only does this reinforce outdated gender-based stereotypes, as Keyes notes, “a trans man who sees a billboard flicker to advertise dresses to him as he approaches is, even if he likes dresses, unlikely to feel particularly good about it.” It would be even worse if AGR systems gain traction in the public sector and are used to control access to public toilets or other gendered spaces, serving to exclude trans people and others who are misclassified by these systems — including in settings like government buildings, hospitals, or vaccination centres where they are seeking access to essential services.”
- [96] See the latest study by AlgorithmWatch and AI Forensics shows that using Large Language Models like Bing Chat as a source of information for deciding how to vote is a very bad idea:
- <https://algorithmwatch.org/en/bing-chat-election-2023/>
- [97] <https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence>
- [98] <https://cdt.org/wp-content/uploads/2023/10/CDT-Comments-on-FEC-Petition-Regarding-Use-of-AI-In-Political-Advertisements-Final.pdf>
- [99] <https://www.paulvallas2023.com/post/vallas-campaign-denounces-fake-deceptive-impersonation-video>
- [100] <https://gcn.ie/cambridge-analytica-gay-rumours-political-candidates/>
- [101] <https://www.tiktok.com/discover/Denis-mukwege-homosexuality>
- [102] <https://factcheck.afp.com/nigerias-opposition-presidential-candidate-atiku-abubakar-has-not-received-lgbt-support>
- [103] <https://factcheck.afp.com/doc.afp.com.33JU3T3>
- [104] <https://factcheck.afp.com/doc.afp.com.32JY8PP-1>
- [105] <https://factcheck.afp.com/doc.afp.com.33AN8U7>
- [106] <https://www.barrons.com/podcasts/barrons-live/marketwatch-ai-and-the-2024-election-how-to-protect-yourself/fc941f84-0fef-4406-b2bc-2c300532d07f?page=1&>
- [107] Interview carried out with Jeremy Safari, Executive Director of Rainbow Sunrise Mapambazuko in Bukavu, DRC.

[e[108] For example, in the US: “Voter suppression tactics frequently target historically disenfranchised communities, including communities of colour, low-income communities, and immigrant communities. During Alabama’s U.S. Senate special election in 2017, residents of Jefferson County — where the largest city, Birmingham, is predominantly African American — received text messages with false information about polling site changes. And on Election Day in 2010, Maryland gubernatorial candidate Bob Ehrlich’s campaign manager targeted African American households with robocalls claiming that Governor Martin O’Malley had already been reelected, implying that his supporters could stay home instead of voting. In 2018, many social media accounts posted false voting information, including instructions to vote by text and claims that voters of one party were required to vote the day after Election Day.”

[109] <https://www.hrc.org/press-releases/new-report-lgbtq-voters-becoming-one-of-the-fastest-growing-voting-blocs-in-the-country-projected-to-represent-nearly-one-fifth-of-voters-by-2040-and-fundamentally-reshape-american-electoral-landscape>

[110] <https://www.concordia.ca/research/migs/resources/rwanda-radio-transcripts.html>

[111] <https://international-review.icrc.org/articles/protecting-civilians-from-disinformation-during-armed-conflict-914>

[112] <https://international-review.icrc.org/articles/protecting-civilians-from-disinformation-during-armed-conflict-914>

[113] <https://www.nytimes.com/2021/05/14/technology/israel-palestine-misinformation-lies-social-media.html>

[114] https://stratcomcoe.org/pdfjs/?file=/cuploads/pfiles/web_stratcom_coe_iran_proxy_war_against_yemen_13-02-2020.pdf?zoom=page-fit

[115] https://www.rand.org/pubs/research_reports/RR4373z2.html

[116] <https://www.ohchr.org/en/press-releases/2022/10/lgbt-persons-must-be-included-un-peace-and-security-agenda-un-expert>

[117] <https://www.corteidh.or.cr/tablas/r39345.pdf>

[118] <https://www.gmfus.org/helping-ukraines-lgbt-community-wartime>

[119] We have not found examples of this online yet, but such interventions are something to look out for considering the targeting of LGBTIQ+ populations in conflict situations.

[120] <https://civiliansinconflict.org/publications/research/when-words-become-weapons-the-unprecedented-risks-to-civilians-from-the-spread-of-disinformation-in-ukraine/>

[121] https://outrightinternational.org/sites/default/files/2023-02/LGBTQLivesConflictCrisis_0.pdf

[122] https://www.eeas.europa.eu/eeas/fimi-targeting-lgbtqi-people_en

[123] <https://factcheck.afp.com/doc.afp.com.32JY8PP-1>

[124] <https://ca.news.yahoo.com/gaza-photo-israeli-soldier-raising-195809274.html>

[125] <https://factcheck.afp.com/post-does-not-depict-palestinian-violence-against-queer-community>

[126] <https://edmo.eu/2023/11/29/what-russian-disinformation-wants-africa-to-think-about-the-war-in-ukraine-and-more/>

[127] <https://docs.google.com/document/d/1SAfEjfl2KxTSdvibAs7mqPUI7WPECVUkaog3ZwV3Z1c/edit>

[128] https://edps.europa.eu/system/files/2021-02/21-02-10-opinion_on_digital_services_act_en.pdf

[129] https://edps.europa.eu/system/files/2021-02/21-02-10-opinion_on_digital_services_act_en.pdf

[130] <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation>

[131] https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence?at_campaign

[132] <https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>

[133] Forbidden Colours will provide a full analysis of the impact of the EU AI act once the legislation has been adopted.

[134] Google’s AI principles include a commitment to not design or deploy “technologies whose purpose contravenes widely accepted principles of international law and human rights”.

[135] Microsoft’s Responsible Innovation Best Practices Toolkit and responsible AI training modules

[136] <https://ai.meta.com/responsible-ai/>

[137] <https://www.washingtonpost.com/technology/2023/03/30/tech-companies-cut-ai-ethics/>

[138] <https://www.theverge.com/2023/11/18/23966980/meta-disbanded-responsible-ai-team-artificial-intelligence>

[139] For example: <https://www.business-humanrights.org/en/latest-news/generative-ai-complicates-accountability-for-online-child-sexual-exploitation/>

[140] On that aspect, see the latest civil society campaign

[141] <https://slate.com/technology/2020/08/facebook-twitter-youtube-misinformation-virality-speed-bump.html>

This report is part of the



The United4Diversity initiative aims at gathering a diverse group of stakeholders around the shared mission of strengthening democratic resilience and ensuring the full access to human rights for all people in Europe. Led by Forbidden Colours, the initiative aims to establish a networking platform, coordinate efforts among participants, and raise awareness about the interlinkage between safeguarding the rights of LGBTIQ+ people and democratic resilience.

This report contributes to the research and awareness-raising initiatives under the United4Diversity initiative.

More information at www.united4diversity.eu

This report was published in January 2024 in Brussels, Belgium.

~~forbidden~~ colours

**Delivering
human rights & democracy
for LGBTIQ+ people in Europe**

forbidden-colours.com