
ARTICLE 19
Red Lines for
AI:
Resisting surveillance,
reclaiming privacy
2024

ARTICLE 19 is an international think–do organisation that propels the freedom of expression movement locally and globally to ensure all people realise the power of their voices.

Together with our partners, we develop cutting-edge research and legal and policy analysis to drive change worldwide, lead work on the frontlines of expression through our nine regional hubs across the globe, and propel change by sparking innovation in the global freedom of expression movement. We do this by working on five key themes: promoting media independence, increasing access to information, protecting journalists, expanding civic space, and placing human rights at the heart of developing digital spaces.

info@article19.org
www.article19.org
@article19org
@article19
facebook.com/article19org

© ARTICLE 19, 2024

This work is provided under the Creative Commons Attribution-NonCommercialShareAlike 4.0 license.

You are free to copy, distribute, and display this work and to make derivative works, provided you:

1) give credit to ARTICLE 19;
2) do not use this work for commercial purposes;
3) distribute any works derived from this publication under a license identical to this one.

To access the full legal text of this license, please visit:
https://creativecommons.org/licenses/by-nc-sa/4.0/

ARTICLE 19 would appreciate receiving a copy of any materials in which information from this report is used. ARTICLE 19 bears the sole responsibility for the content of the document.

Design and typesetting by Sharon Leese.

Acknowledgements

This report has benefitted from the advice and insights of many experts from civil society, academia, and international organisations. We would like to thank the participants of online meetings as well as those who gave us time for individual conversions to discuss their views and reflections on the drawing of AI red lines. We would also like to thank all the colleagues who reviewed earlier drafts of the report.

Executive summary

This report identifies the contextual, social, and political factors that may shape decisions on whether to call for red lines on the use of artificial intelligence (AI) or whether to focus on other routes to challenging use cases, including through engagement with companies and the use of regulatory and judicial forums. The report finds that these are not necessarily either/or choices, in that campaigns to adopt red lines for biometrics are likely to benefit from parallel challenges to technological applications in other forums. Moreover, as social and political contexts shift, space may open for the pursuit of red lines on AI, including as a result of a moratorium on the use of particular technologies, successful outcomes in court, or the voluntary abandonment or temporary pause of the design, development, or sale of technologies by companies.

This report aims to equip civil society organisations with the information and considerations necessary to successfully pursue red lines in the specific contexts in which they work. It explores the current state of play, reflects on the current obstacles to having red lines adopted by governments, and identifies key factors for civil society organisations to consider at the time of determining their own strategies to establish red lines.

The report first identifies the three main reasons often given for banning specific technologies or use cases which underpin calls for red lines: inaccuracies in performance; inherent, unnecessary, or disproportionate risks to human rights which cannot be mitigated; and the exacerbation of power imbalances between institutions using face recognition and individuals.

Second, the report maps where red lines on the use of AI have been secured, highlighting the variances in the scope and nature of such red lines in terms of technological coverage, actors, and geography.

Third, the report offers recommendations for civil society and other actors to consider when deciding on whether and how to pursue red lines for biometrics, including through explicit advocacy campaigns, policy, and law reform or litigation.

Contents

Introduction

The human rights impact of artificial intelligence (AI) technologies and applications[1] has become increasingly well understood around the world. From biometric surveillance[2] to generative AI products like ChatGPT, AI technologies receive significant funding from powerful actors, while at the same time garnering attention for potential areas of misuse and discrimination, exploitative supply chains, and environmental harm.[3]

For some of these technologies, civil society organisations have called for bans, or 'AI red lines', where they do not comply with human rights standards. Civil society calls for AI red lines also arise because the intended applications present risks that are too high for a society to take, and very often disproportionately impact historically marginalised groups. For instance, civil society actors have called for bans on the use of certain AI-enabled technologies in migration contexts, including, but not limited to a ban on 'remote biometric identification and categorization in public spaces, including in border and migration control settings'. Some civil society organisations have adopted explicit advocacy campaigns calling for bans on the use of facial recognition technology.[4]

In other instances, even if civil society and other actors seek the banning of a technology, they have not explicitly called for the drawing of an AI red line but have rather focused on securing a moratorium – or temporary cessation of the deployment of a technology – to give space for debate and analysis of its societal impact, whether any potential harm can be mitigated, and the adequacy and effectiveness of safeguards in place.

An AI red line would most clearly be drawn through a legislative act. Already some technologies have been subject to a legislative ban at the local[5] or regional levels. Elsewhere, specific use cases have come under challenge through regulatory and judicial bodies and through engagement with the companies designing, developing, and selling certain applications. Although these efforts cannot be characterised directly as calls for AI red lines, much like moratoriums, they form part of wider initiatives aimed at defining the parameters of if and how AI applications should be designed, developed, and deployed and can build momentum to the formal adoption of bans or AI red lines through legislation.

More recently, civil society efforts to secure AI red lines have expanded to biometric technologies more broadly as well as to other forms of technologies and use cases. In recent months, there have been a few calls for banning the development and use of large language models; however, these efforts have been relatively less mature.

This report is designed to support this advocacy. It first takes a preliminary look at how the pursuit of AI red lines are approached by civil society stakeholders, in particular against the use of facial recognition technologies by law enforcement as an area in which we have the most robust data. Informed by expert group meetings organised by ARTICLE 19 and allied organisations, as well as by interviews with civil society organisations in different countries, this report aims to equip civil society organisations with the information and considerations necessary to successfully pursue red lines in the specific contexts in which they work. While it is aimed for global coverage, and indeed engages with civil society efforts in different countries, the report pays significant attention to developments in the US. This is because the US is one of the few countries, to date, where AI red lines have been achieved, particularly at the city level. Further, public appeals to major tech companies have resulted in commitments to cease the sale of certain biometric technologies until the introduction of dedicated regulation in the US.

Importantly, the report identifies considerations that civil society might want to take into account if they decide to pursue red lines, including factors that may point against such explicit calls even if that is the ultimate goal.

We hope that the report will be a starting point for discussion, debate, and analysis rather than offering definitive and conclusive findings on the establishment of AI red lines. We also invite further research and dialogue on this topic, including through the building of a repository of examples and experiences of civil society efforts to establish AI red lines.

AI red lines


int main(int argc, char **argv) {
vector<char> acc;
char ch;
ifstream infile(argv[1]);
argv) {

Main justifications for the establishment of AI red lines

Not all actors use the language of AI red lines. Some simply call for the banning of the design, development, or deployment of a technology. Others argue that the use of a technology does not comply with international human rights standards. Regardless of the language employed, the following main arguments may be employed to argue for the cessation in the development and use of a technology, including when explicitly calling for an AI red line.

Incompatibility with international human rights standards

A fundamental argument employed to justify a ban, or AI red line, relates to the human rights their design, development, and deployment put at risk either inherently or in a way that is unnecessary and disproportionate. Therefore, it fails to comply with international human rights standards and norms.

The risks to human rights constitute a central justification where bans have been adopted.[6] These risks are underscored in campaigns to ban facial recognition and remote biometric recognition technologies.[7] The campaigns point to risks to the rights such as to privacy, freedom of assembly and association, equality and non-discrimination as well as the right to liberty and due process, where the use of these technologies results in wrongful arrest.[8] They also emphasise the propensity for the use of these technologies to accentuate existing forms of discrimination and inequality in society, thus disproportionately impacting minorities and groups in positions of vulnerability,[9] particularly within cities.[10] Further concerns arise about the targeting of journalists, human rights defenders, and groups that are put in vulnerable positions by the actions of states, such as people on the move, for surveillance.[11]

Calls for bans on emotion recognition technologies also point out the violation of multiple human rights, including human dignity and unacceptable intrusions into an individual's private mental life. They also argue that these technologies are based on pseudoscientific claims of inferring people's inner emotional states, which legitimises discredited scientific arguments and heralds the resurgence of physiognomic thought. Similar arguments have been made in relation to polygraphs.

Additionally, arguments against the use of AI-based risk assessment systems in the migration context highlight how the use of such systems would violate the right to non-discrimination, create risks of indirect discrimination, and violate the rights to privacy, data protection, and fair procedure. These arguments are not made solely on the basis of the technical system alone, but rather on how technologies operate within institutions with their own unique historical context.

Inequality and power imbalances between institutions and individuals

Civil society organisations emphasise the imbalances of power resulting from the harnessing of facial recognition and other forms of biometric technologies, particularly by companies and states. For example, they argue that facial recognition can contribute to moves towards authoritarian governance and new ways to control individuals in society.

In multiple jurisdictions, from the UK to India, decisions on whether or not to develop these technologies are often left out of public discourse, with civil society only finding out about deployments post-fact or through newspaper reports.[12] Our interviews surfaced another pertinent point – concerns about the impact on democracy are often only expressed with regard to public sector use of such technologies. For instance, the discourse on the use of biometric technologies in the workplace primarily focuses on issues of privacy and/or effectiveness, but rarely, if ever, on structural harm caused by these technologies or of concentrating power in the hands of those who control capital to be wielded over those who provide labour. This is particularly important to note given the fact that intrusive technologies often reach the public sector after having been introduced and normalised through commercial spaces, such as the workplace, meaning that they are important sites of focus when constructing calls for bans or moratoriums.[13]

Inaccuracies in the technology

Some civil society actors have highlighted that claims about what the technology is capable of can be overstated and the technology may produce high error rates with serious consequences for the protection of human rights. For example, emotion recognition technologies have been characterised as a form of pseudoscience that legitimises the widely discredited Basic Emotion Theory (BET).[14] The very existence of emotion recognition is what poses the threat to human rights, and by definition, these systems do not work – their accuracy is thus a non-starter.

For facial recognition technologies, studies demonstrate that commercial facial recognition software have significant gender and racial biases. These inaccuracies carry serious consequences for human rights, including discrimination, misidentification, wrongful arrest, and detention.[15]

While the inaccuracies of these technologies are a critical argument against their use, civil society actors have cautioned against solely relying on this argument to avoid the implication that the reduction or elimination of error alone would fully address the human rights challenges associated with their use.[16] For example, it has been argued that 'face surveillance is dangerous when it works and when it doesn't, and there's a lot of reason to believe that the technology doesn't work very well'.[17] When technical arguments are made, they are usually quickly followed up with a reminder that accuracy will get better over time. However, while improving accuracy may address some issues with these systems, 'this will ultimately only perfect them as instruments of surveillance and make them more effective at undermining our rights'.[18]

AI red lines


```
int main(int argc, char **argv) {
   vector<char> acc;
    char ch;
   ifstream infile(argv[1]);
```

The current state of play of AI red lines

While civil society actors increasingly call for the drawing of AI red lines, these campaigns remain in early stages of development, as does the adoption of dedicated legislation on AI technologies in general.

Some use cases have been banned, most notably in US cities. This can be attributed to civil society advocacy to a considerable extent.[19] Elsewhere, even the most robust, coordinated and mature civil society advocacy has not resulted in red lines, most recently in context of the EU AI Act.[20] In other cases, moratoriums have been adopted by states and voluntarily by companies through public, reputational, investor, and employee pressure. Regulators and courts have also found specific use cases in violation of constitutional, data protection, or human rights law. These are also largely the result of civil society advocacy.[21] While these are not formally AI red lines, they may pave the way for the later legislative banning of categories of technologies or use cases.

This section illustrates the types of red lines that have emerged across jurisdictions and the measures that fall short of, but could potentially lead to, red lines. The types of red lines are organised based on the entity putting in place the red line, that is, legislators, private companies, or the judiciary.

AI red lines in legislation

To date, very few legislators have enacted AI red lines. Dedicated legislation on AI technologies and calls for legislative prohibitions and bans remain in their infancy, meaning that more AI red lines may be drawn within legislation on AI technologies in the future. It is therefore important to understand the types of AI red lines that have been successfully adopted to date, as well as their limitations, to avoid similar challenges in the future.

The scope of adopted and proposed legislative AI red lines

The US provides the majority of the current examples where most bans secured have been at the city level. The first of these bans was adopted by the San Francisco Board of Supervisors in 2019,[22] followed by other US cities, including Berkeley, CA, Cambridge, MA, and Boston, MA.

The nature of these bans is not uniform and they differ both in terms of scope and exemptions. Some bans not only prohibit city officials and departments from obtaining, retaining, accessing, or using facial recognition technology or face surveillance systems but also prohibit them from requesting, possessing, selling, or evaluating such technology; using information obtained from such technologies; or entering into agreements or issuing permits to third parties to use such technologies.[23] The bans therefore vary with regard to factors such as whether they are limited to direct acquisition or use of facial recognition technologies, or they also enable third parties to use the technologies and the state to benefit from any resulting information. Some bans also specify whether the ban only covers the financial purchase of such technologies or whether donations are also within the scope of the ban. This is an important detail as in some instances states acquire technologies through donations or gifts by technology companies which can, in some jurisdictions, enable the bypassing of public scrutiny of the legality of the introduction of a technology, even if on a trial basis, into the public sector.

By contrast to the US city bans, the EU Artificial Intelligence Act takes a risk-based approach and provides a list of applications deemed 'unacceptable'. In June 2023, the EU Parliament voted to ban multiple public mass surveillance applications of biometric systems.[24] However, these safeguards were significantly watered down during the trilogue negotiations, resulting in a final text that contains significant loopholes.

The impact of extensive exemptions within adopted and proposed legislative AI red lines

While some legislative AI red lines have been achieved, many are more limited than they initially appear due to the inclusion of exemptions, with some more expansive than others. For example, within the US city bans, exemptions have existed for personal use,[25] public communications exemptions,[26] redaction software exemptions,[27] exemptions for officials' inadvertent or unintentional receipt, retention of, access of, or use of any information obtained from face recognition technology,[28] exemptions for use in relation to a criminal investigation,[29] or missing persons exemptions.[30] The exemptions are also provided in the EU Artificial Intelligence Act.[31]

The inclusion of exemptions 'can swallow any rule'.[32] They undermine the significance of a legislative ban and can also result in the technology being used in some of the highest risk situations for human rights, such as criminal investigations. These exemptions may, in effect, dismantle the ban and enable the use of technology in contexts in which there are already serious concerns about discrimination, such as the investigations of crime, and that may have some of the most serious consequences for human rights, such as arrest and detention. This greatly depends on who decides when the exemption applies, how much discretion is allowed in these circumstances, and what safeguards apply once the exception is granted. Importantly, if regulation is put in place, and if it is permissive, it can also have the effect of encouraging and legitimising such use.

Roll-back of some legislative AI red lines

In addition to exemptions, some US city bans have been reversed which underscores the potential fragility of some legislative AI red lines and the need to continue working to maintain them.[33]

Measures which fall short of, but may lead to, the adoption of AI red lines

While few legislative AI red lines have been adopted, civil society organisations and other actors, including employees of major tech companies, have pursued other avenues to limit the design, development, and deployment of AI technologies and which carry the potential to lead to AI red lines, now and in the future. To date, most initiatives have been limited in geographic scope, with many connected to the US markets.

Adoption of moratoriums on the use of AI technologies

In some cases, attempts have been made to enact a moratorium on particular use cases in the US[34] and in Italy.[35] A moratorium provides the space to consider the potential harm presented by a particular technology or use case and to assess whether it should be subject to a legislative AI red line or whether better safeguards are required to mitigate the potential harm. However, the nature of moratoriums can also mirror challenges presented with explicit AI red lines, particularly with regard to exemptions.[36]

Voluntary commitments by private sector actors to stop developing or selling AI technologies

In some cases, technology companies have voluntarily committed to stop developing or selling AI technologies to state agencies until regulation is adopted in the US, implicitly limiting the reach of the moratorium to the US rather than globally.

The motivations for such voluntary commitments may be manifold or differ depending on the specific technology, context, or use case, but may include companies wishing to manage their public image following controversial and harmful uses of certain technologies; pressure from employees or shareholders; or a focus on self-regulation as a means to avoid regulation by state(s). In theory, such efforts may lead to the adoption of legislative AI red lines but could also result in more permissive uses of AI technologies, thereby requiring critical engagement by civil society.

These voluntary commitments fall into one of three distinct categories. The first category relates to the adoption of general principles on when a company will desist from the design, development, and deployment of AI technologies.[37] The second category relates to commitments by companies not to sell or transfer their technologies to state entities.[38] Where companies introduced moratoriums, they typically explained their introduction as a means for the US Congress to adopt regulation on the use of such technologies, particularly by law enforcement, thus potentially opening up a pathway for the adoption of a legislative AI red line.[39] At the time of writing, however, no such regulation has been adopted.

The third category of voluntary commitments, and possibly the least developed, is the prohibition of B2B sale/transfer of certain technologies to private actors. In contrast to state procurement of face recognition, there has been relatively little focus from companies to monitor downstream uses, that is, how these technologies are sold by private companies to other private companies.[40]

These initiatives are critical to engage with as they can potentially stop the development and deployment of AI technologies which are harmful to human rights quickly and at source, thus also preventing state actors from acquiring them. However, they may only be temporary and can be as quickly undone.[41] Further, as discussed later, despite the global reach of major technology companies, to date these voluntary

commitments appear restricted to the US, and thus much more work is required to widen their scope and reach. Equally, where companies connect their commitments to formal regulation they, theoretically at least, create impetus for the adoption of legislation of AI technologies, although how such legislation is drafted and the political context surrounding it will shape whether such legislation contains AI red lines or creates a permissive environment for technological deployment. While these private initiatives may push for the adoption of these legislations, these initiatives risk leaving such important decisions to profit-orientated business strategies.

Pursuit of constraints on specific use cases of AI technologies through challenges to regulatory and judicial bodies

Finally, a small body of decisions by judicial (for example, in the UK[42] or Brazil[43]) and independent regulatory bodies (for example, in the UK[44]) assessing the compatibility of specific technological deployment with constitutional, data protection, and human rights principles has started to emerge. While these decisions are limited in scope to the facts of the particular complaint, they have the potential to build momentum to securing a wider legislative AI red line. Depending on the political context, challenges to state and business use of AI-enabled technologies before judicial and regulatory bodies may complement efforts to secure legislative AI red lines or may present a more strategic route to restricting technological deployment, depending on the circumstances.

Accordingly, while these measures do not constitute formal AI red lines and remain relatively restricted in geographical scope, they illustrate the paths that can be pursued prior to, or alongside, campaigns to adopt AI red lines. These measures carry the potential to build support for legislative AI red lines as well as engage a wider community of actors calling for such measures, such as employees of major technology companies.

At the same time, in states with low institutional, regulatory, policy-making, and enforcement capacity and in which simple solutions to addressing complex issues at scale are sought, solutions offered by the private sector and tech companies oversimplify the problem at hand, and frame it in a way that presents the use of technology as the magical answer (also called a tech-solutionist approach) and without recognition of the difficulties entailed in working with data.[45] In addition to the pathway where companies provide technology to states at no/low costs to expand their market potential, it is important to note that the private sector often creates demand for its products and services. By defining the problem, and marketing their products as solutions to the problem, they are also then shaping the governance models adopted.[46]

This is perhaps most pronounced in the case of 'smart cities', where problems of governance and public service delivery are framed in terms of inefficiency and lack of data and information sharing by private companies. These governance problems are therefore 'solved' through the use of technologies that facilitate seamless data collection, storage, and sharing.[47] But technology applied to complex social problems solves very little and benefits those who already enjoy a certain amount of privilege. The fundamental issues of historical discrimination leading to discrepancies in access to resources and government services, underpaid government staff resorting to corruption and under-performing at their jobs, and the 'inefficiency' of providing care and support to communities that have been made vulnerable by societal and political actions cannot be made to disappear into thin air by technology.

AI red lines

Our recommendations for strategy development and positioning on AI red lines

Securing AI red lines and measures which can create a pathway to them is a challenging and time-consuming process. This is made even more difficult by the lack of transparency that often attends state and business practices. States and businesses have the responsibility to disclose their plans to deploy AI technologies, to carry out and publish meaningful human rights impact assessments, including the safeguards they propose to put in place to mitigate any potential harm. Such an approach would allow for public scrutiny and debate about whether and how AI technologies should be deployed ahead of time.

At the same time, civil society, academics, and journalists should play an active role to investigate and challenge the deployment of AI-enabled technologies by states and businesses. Some civil society actors have already created new approaches for investigating government practices where there is little to no government transparency. Novel practices to uncover data and information about the existence, roll-out, and impact of facial recognition technology, for example, are crucial to effectively push for red lines in the long run.

In this section, we put forward recommendations for strategic considerations and actions to take into account when considering the nature, scope, and avenue of AI red lines to pursue.

Political and local realities may therefore influence decisions on AI red lines, particularly in the current absence of international prohibitions, and limited legal interventions at the regional level, on the design, development, and deployment of particular technologies.

Recommendation 1:
Incorporate demands for state accountability and transparency as part of AI red lines strategies

As highlighted earlier, several of the bans – as well as moratoriums – were introduced after the AI-enabled technologies had already been deployed and after individuals and groups had already been harmed by the technology. Technology is typically deployed without close or public scrutiny of the potential human rights impacts of the technology and assessment of whether such risks can be mitigated. This highlights the lack of transparency from the state about the types of technologies, use cases, and public/private partnerships under consideration, which makes it nearly impossible to secure bans before deployment. This includes situations where companies may be testing or trialling new technologies,[48] or where companies 'donate' technologies to states,[49] particularly in less regulated spaces.[50]

The lack of transparency and dedicated processes for scrutinising AI-enabled technologies prior to deployment also shifts the burden from states and companies wishing to roll out the technology to civil society actors, academics, and investigative journalists who then have to both uncover their use and build a persuasive narrative about the harms posed by these technologies. Once deployed, such actors are likely to face greater hurdles in securing red lines, particularly where states have invested in AI systems.

To the extent the status quo continues, civil society should continue to adopt reactive approaches as and when information is revealed or discovered about technologies already deployed. It is equally important to develop strategies to address the opacity of states and private actors regarding procurement, public–private partnerships, and actual deployment of AI-enabled technologies. Demands for transparency should focus, in particular, on the following:

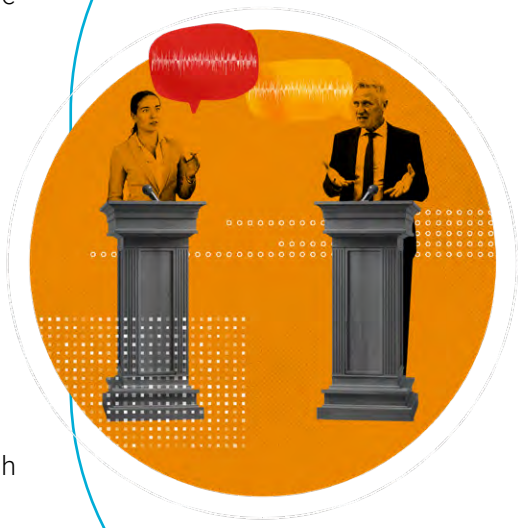
1. **Push for the adoption of a legal obligation for state actors to issue public notices of any plans to employ AI-enabled technologies.** This obligation would enable public scrutiny and debate on proposals to deploy AI-enabled technologies before the decision to deploy has been made. Civil society should advocate to oblige state agencies to provide clear and detailed information on their reasons for seeking to deploy an AI-enabled technology. States are also obliged to assess whether the technology meets the test of legality, necessity, and proportionality to the aim pursued; and whether there are the safeguards to prevent human rights harm. To enable meaningful public scrutiny, the obligation should include clear procedures for input to be provided and considered, and proposals to be modified or abandoned entirely.

2. **Advocate for public disclosure and scrutiny of state procurement processes, from initial conceptualisation to consideration and decision-making on specific technological models and providers.** We believe that the focus on procurement is important due to the insights that can be gleaned into how and why states deem technology to be a promising investment in pursuit of governance goals. Procurement, and the investment financing it, should not only include situations where the state purchases technology, but also where private actors approach the state to donate or offer access to technology without charge or as a 'pilot' project. A granular understanding of the procurement process can clarify how assumptions about the utility of a technology and partnerships are cemented in policies and approaches to using AI-enabled technologies and the extent to which human rights considerations are considered, including through human rights impact assessments which should be conducted at each stage of the procurement process.

3. **Submit access to information requests about the use of technologies in government departments, and/or access documentation on company projects (if possible).** In states that have adopted access to information laws, requests under these laws have been instrumental in gaining insight, or revealing the lack of existing information, about state use of technologies, particularly in cases where technological use has been made publicly known after the fact. Depending on the context, access to information laws can be a useful avenue to pursue greater accountability, not just to understand the priorities and activities of government departments, but also those of the companies contracted by state actors. Importantly, depending on the timing and scope of the requests, access to information processes can also pave the way for the disclosure of details of the potential donation or procurement of specific technologies and the nature of potential public/private partnerships before deployment, creating time for public debate, including on the matter of setting AI red lines.[51]

Recommendation 2:
Develop a clear and multidimensional narrative on the need for AI red lines

We believe that such narratives of red lines should **demonstrate risks posed by the technology itself along with the existing inequalities**, power imbalances, and threats to human rights in the contexts into which these technologies are interjected. They should also show how the context can result in technology worsening or introducing new threats and highlight the inadequacy of the current legal framework in protecting against the harms caused by such technologies. At the same time, it is also crucial to put in place strong arguments that disprove tech-solutionist narratives that spur investment, development, and use of such technologies. This is best done through collective brainstorming and strategic coalition work within civil society.



The previous section points to the importance of identifying the types of problems that states and private actors claim they are seeking to address in adopting technologies. Civil society should **interrogate the justification and evidence** for such claims. A critical analysis of the capacity of AI-enabled technologies to 'solve' often complex social problems can reveal the lack of sufficient evidence and the overpromising of the capabilities of specific technologies. It can also help generate a wider discussion and debate on the types of approaches and resources needed to address the problem concerned, such as crime rates.[52] Additionally, civil society should also develop strategies to enhance the political, democratic, economic, legal, and social conditions necessary to prevent abuse and the exacerbation of existing inequalities through the use of these technologies.

Furthermore, civil society should **avoid agreements on red lines that contain wide exceptions to permit the use of AI** for 'national security', 'public safety', and other broadly defined use cases. These exemptions could also have public support. For example, the public may object to the use of biometric technologies for commercial purposes,[53] but agree to them when these technologies are presented as a means to protect people from crime.[54] This is problematic because, typically, measures justified on grounds of public safety often result in some of the worst violations of people's human rights.[55] For instance, while states may justify the use of facial recognition to locate a missing child or counter terrorist attacks, in reality the portrayal of such uses as 'exceptional' belies the extent of the mass surveillance practices that such exceptional cases would require.[56] Such use cases do not meet the tests of necessity and proportionality, even if the purpose is legitimate: this is also not the least intrusive way to solve this problem, nor the only way.

Civil society should strive for a much greater understanding of how the technology would work in exempted cases to determine whether such exemptions are justified. In this regard, civil society should advocate for **the onus to be placed on those arguing for exemptions**. They should specifically prove that:

- The restrictions of rights meet the narrowly constructed test of legality, necessity, and proportionality and that the technology is necessary to achieve the aim the state is pursuing.
- A moratorium or ban would adversely impact public safety and no less intrusive means exist to meet the same objectives.

We further recommend that **narratives on the need for AI red lines should be developed through an inclusive and participatory process**. The process of bringing together multiple stakeholders beyond human rights or digital rights groups,[57] including the public, can collectively diagnose the problem posed by the use of AI-enabled technologies and build dialogue and consensus on AI red lines.[58] This is particularly important in light of the promotion of AI-enabled technologies to address problems such as high crime rates or public safety.[59] A broader support base for AI red lines advocacy, unified civil society demands, and stronger networks increase the vectors as well as the range of actors that can exert influence.

Civil society should also **engage actors that can exert influence over states and companies** and bring them into the process of problem-diagnosis. For example, some civil society organisations have worked closely with investors and independent researchers as critical stakeholders in debunking myths and identifying the potential harms posed by new and emerging digital technologies and in influencing companies in the policies they adopt in response.[60]

Recommendation 3:
Carefully assess the potential for pitfalls or blowback when identifying advocacy opportunities

Pursuing red lines may constitute one (important) part of a wider strategy on minimising harms arising out of the sale or use of AI. When technologies are proposed as solutions to complex social issues, the narrative of 'innovation' and 'modernisation' can be appealing to stakeholders in power. In such contexts, some civil society actors have dismissed the adoption of explicit campaigns for AI red lines as being politically unfeasible or have assessed that such calls could lead to the adoption of permissive legislation, especially where states have already invested in intrusive AI-enabled technologies.

In such situations, civil society can consider **the pursuit of AI red lines through litigation** using existing laws and focusing on data protection, freedom of expression, or non-discrimination and the necessity and proportionality of the technology.[61] These are important building blocks towards securing red lines. Highlighting the weaknesses in existing regulatory frameworks can be part of the wider diagnosis of the problem and point to the extent of the potential harm if biometric technologies were adopted. Furthermore, there may be risks in explicitly calling for red lines, and so the choice of best route to take must also consider this reality.

At the same time, civil society actors should keep in mind that challenging the necessity and proportionality of biometric technologies in courts can also lead to confirmation of the sufficiency of existing law (for example, on data protection), even if not specific to these technologies. Equally, these are not static positions and so can change depending on the political climate. Identification of key stakeholders can ultimately influence the regulatory and governance environment. When possible, civil society should take all these strategic factors into account when identifying the litigation forum.

Recommendation 4:
Determine the type and scope of the AI red lines

Type of AI red lines

Civil society actors seeking AI red lines may have to decide on the policy position to adopt – from support for a moratorium or a ban (full prohibition). For certain technologies or use cases, the introduction of safeguards will not be able to lessen the risks to people's human rights and full prohibition is an ideal goal. When deciding whether to support moratorium, we suggest they consider the following issues:

1. **A moratorium may 'buy more time' for discussion and debate.** This is particularly the case where states and companies emphasise the importance of innovation or the need for biometric technologies on grounds of public safety. Securing a moratorium, especially prior to deployment of such technologies, can create space to engage with such arguments, to demonstrate the human rights impact of biometric technologies. A moratorium also provides an opportunity to generate greater awareness and consensus-building on the necessity and proportionality of such technologies, such as whether they are capable of protecting public safety and whether alternatives exist, including in non-technological form.

2. **An explicit pursuit of full prohibition may adversely impact the civil society's ability to participate in and shape the governance and regulatory landscape.** For example, some argue that the state will use the technology in any case, and the calls for a ban can be 'the end of the conversation', as one civil society colleague put it.[62] Focusing on a moratorium may therefore reflect a strategic choice when the democratic and political fabric of a given local context does not entertain the idea of a ban. In this regard, the pursuit of a moratorium on the use of AI technologies may present a strategic first step to open up conversations on AI red lines.

Scope of AI red lines

The scope of AI red lines can vary and may have to evolve and adjust over time. Civil society actors should consider that each of those presents their own merits and challenges:

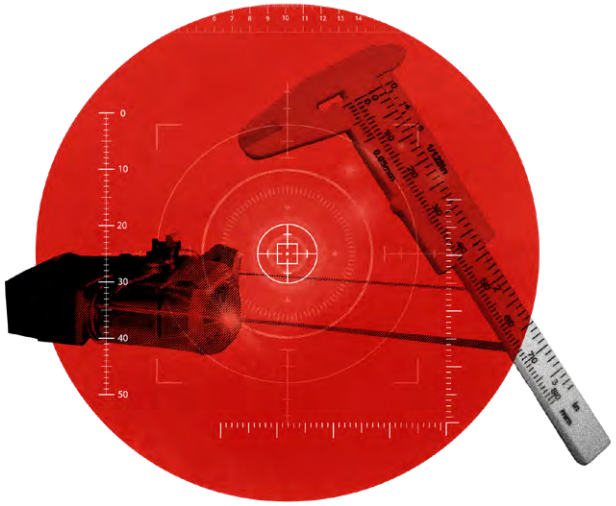
1. **General principles on the prohibition of AI-enabled technologies.** General bans apply, at least in theory, to current and future forms of technology and use cases. We find that this could prevent the need for developing legislation for each form of technology or use case and potentially overlooking harmful applications of AI-enabled technologies. However, an overarching prohibition can present challenges in practice. To be effective, it requires the articulation of detailed enough principles that would enable both state and private actors to identify use cases falling under the general prohibition and a mechanism for determining where states and private actors had failed to comply with the prohibition.[63] Furthermore, arriving at a general prohibition of particular technologies may be challenging where state and private actors argue that these technologies also have 'uncontroversial', and even productive, uses.[64]

2. Intermediate step: **an overall ban of particular types of technologies based on specific forms of processing data and types of data**, without specifying the use case or actor (intermediate solution). In our experience, this can avoid a technology-by-technology approach, including where limited by actor, which can overlook commonalities in the risks posed by different technologies that may also merit inclusion within a ban, even if they have not yet been deployed or proposed for use.[65] This type of approach can potentially avoid situations where prohibition of the use of certain biometric technologies by law enforcement, for example, has led to criticisms for its exclusion of the use of the same types of technologies in other contexts such as border governance.[66]

3. At the same time, We recognise that a prohibition of particular types of technologies may result in a focus on technologies that already exist and thus fails to be general and forward-looking enough to encompass future forms of potentially harmful technologies. **Where list-based prohibitions are pursued, civil society actors should present the list as one that evolves over and with time**, unlike the EU AI Act, which includes a closed list. However, an open list gives the possibility of hard-won red lines to be reversed in the future – and is thus a balancing act that must be deliberated on in each given context.

4. **A targeted approach to AI red lines.** This approach can enable the development of a clear narrative on the human rights risks posed by the design, development, or deployment of specific technologies by particular actors and allow for the incremental building of support for their prohibition.[67] Where prohibitions are secured for some of the most well-known and sharp end-uses of AI technologies, this can build the basis for future expansion of red lines. At the same time, targeted calls for AI red lines and narrow formulations may imply that only certain forms of biometric technologies present risks, or only when used by specific actors. This approach may also overlook the interplay between state actors and between the state and the private sector. It is therefore difficult to have clear demarcations of bans confined to use cases or actors given the blurred lines between the public and private sector and the potential for purpose and function creep.

The challenges associated with each approach underscore the fact that civil society actors should clearly explain the reasons behind the call for a prohibition or moratorium. As our experience shows,[68] civil society should also pay close attention to the process by which calls for AI red lines are developed. This requires attention not only at the initial stages of formulation but also as campaigns are implemented, as opportunities may arise for coalitions and connections with actors calling for red lines in other contexts.[69]



Recommendation 5:
Determine the geographic scope of the AI red lines

When planning advocacy on AI red lines, civil society actors should consider the fact that most current determinations of AI red lines are geographically limited.[70] While many factors shape the geographical boundaries of AI red lines – for example, some limitations are inevitable given the specific jurisdiction of local or national legislatures and judiciaries – there are concerns about the impact of securing bans within particular geographical areas rather than globally.

For example, where companies have committed to a moratorium or to move away from the development or sale of specific technologies until states develop dedicated regulatory frameworks, they have been unclear as to whether their commitments are contained to the US or extend to the global market in which they operate. These instances seem like wins at the local level but are actually incredibly costly globally. For instance, Amazon and Microsoft create the illusion of change by voluntarily committing to stop or pause technological sales to state actors. But by confining it to a single jurisdiction, they simply obscure the problem at hand – facial recognition systems built by big tech companies are often trialled first in countries in the majority world, then made available the world over, and eventually become embedded in law enforcement agencies globally. Committing to a moratorium in a single jurisdiction demonstrates companies responding to public or legal pressure, and not fundamentally reckoning with the implications of their products.

In this regard, the risk that multinational tech companies continue to sell facial recognition technologies to law enforcement outside of the US heightens the argument for a complete ban. This argument calls for prohibitions within the jurisdiction at hand, and also prohibits development and cross-border export – which is particularly important in the context of states with weak legal frameworks or poor governance.[71] Localised bans may mean that certain companies can no longer sell their products elsewhere, but this may simply open up the space to other companies from regions without a ban in place.

For these reasons, by recognising the international nature of the supply chain of AI technologies, human rights bodies have called for a moratorium on the export, sale, transfer, use, or servicing of privately developed surveillance tools until a human rights-compliant safeguards regime is in place.[72]

It is therefore critical for civil society to engage with the limitations of bans and moratoriums and to make explicit the ideal scenario: red lines that are applied in one part of the world should ideally be extended globally as human rights deserve equal protection everywhere. A red line that is divorced from a reckoning with the global supply chain of AI technologies is ineffective at best, and misleading at worst. At the time of advocating for red lines, civil society groups should also emphasise that these standards and prohibitions should extend to import and export of products as well, and demand greater disclosures from companies about the entities that they make their products available to.

Beyond advocating for global red lines, civil society actors should take advantage of the fact that the establishment of AI red lines in one locality can have a multiplying effect.[73] Moreover, civil society organisations have sought to overcome jurisdictional limitations by bringing proceedings in multiple jurisdictions.[74]

Conclusions

Civil society's advocacy to secure bans or moratoriums remains in the early stages, particularly as states and regional and international organisations are only beginning to consider formal regulation of AI technologies. As such, the lessons learned from early attempts, successes, and failures to establish AI red lines offer important insights for shaping emerging and future regulation. They are also important for developing other strategies and avenues by which to pursue AI red lines, such as through judicial and regulatory bodies.

While these forums have their own limitations, as discussed in this report, they can also provide opportunities for close analysis of the legality, necessity, and proportionality of technological use cases and thus contribute to the banning of technologies which irreversibly put human rights at risk. Civil society should assess these opportunities, including the risk of legitimising existing use cases or regulatory regimes that have been declared adequate, before proceeding.

Endnotes

[1] For the definition of AI, see ARTICLE 19 and Privacy International (2018) Privacy and Freedom of Expression in the Age of Artificial Intelligence, 6–7.

[2] See, for example, Hill, K. (2020) 'Wrongfully Accused by an Algorithm', *New York Times*, 30 August; Johnson, K. (2023) 'Face Recognition Software Led to His Arrest. It Was Dead Wrong', *Wired*, 20 February.

[3] See, for example, Harwell, D., Tiku, N. and Oremus, W. (2022) 'Stumbling with their words, some people let AI do the talking', *Washington Post*, 10 December; Forbes, (2023) 'Microsoft confirms Its $10 Billion Investment Into ChatGPT, Changing How Microsoft Competes With Google, Apple And Other Tech Giants', 27 January; Russell, M. and Black, J. (2023) 'He's played chess with Peter Thiel, sparred with Elon Musk and once, supposedly, stopped a plane crash: Inside Sam Altman's world, where truth is stranger than fiction', 27 April; Birhane, A. and Raji, D. (2022) 'Chat GPT, Galactica and the Progress Trap', *Wired*, 9 December; McQuillan, D. (2023) 'We come to bury ChatGPT, not to praise it'; Weill, E. (2023) 'You Are Not a Parrot', *Intelligencer*, 1 March.

[4] See, for example, the American Civil Liberties Union (ACLU) of Massachusetts's Press Pause on Face Surveillance, Amnesty International's Ban the scan or EDRi's Reclaim your Face. See also an open letter signed by over 170 organisations calling for the ban on biometric surveillance in June 2021.

[5] San Francisco, CA, banned the use of facial recognition technologies by city agencies in May 2019; Somerville, MA, banned the technology in June 2019; Oakland, CA, banned the technology in July 2019.

[6] For example, the Ordinance adopting the ban on facial recognition technologies in San Francisco highlights that the 'propensity for facial recognition technology to endanger civil rights and civil liberties substantially outweighs its purported benefits, and the technology will exacerbate racial injustice and threaten our ability to live free of continuous government monitoring.' See also Ordinances in Boston, MA, and King County, WA.

[7] Open letter of over 170 organisations.

[8] Open letter of over 170 organisations.

[9] Open letter of over 170 organisations.

[10] Interview 4.

[11] Scassa, T. (2021) 'Privacy in the Precision Economy: The Rise of AI Enabled Workplace Surveillance during the Pandemic', Centre for International Governance Innovation, 8 June; Kaur, H., McDuff D. and Williams, A. C. (2022) 'I Didn't Know I Looked Angry': Characterizing Observed Emotion and Reported Affect at Work', CHI.

[12] See, for example, Privacy International, 'King's Cross has been watching you – and the police helped'; or Ulmer, A. and Siddiqui, Z. (2020) 'India's use of facial recognition tech during protests causes stir', *Reuters*, 17 February.

[13] See, Google, AI Principles 2020 Progress update, 10–11 (providing explanations of how the design of particular technologies were adjusted).

[14] BET posits that it is possible to gauge a person's inner emotional state from their outer facial expression, and such expressions are discrete and uniformly expressed across the world.

[15] See, for example, Johnson, T. L. and Johanson, N. (2023) 'Police Facial Recognition Technology Can't Tell Black People Apart', Scientific American, 18 May, which found that the foundational takeaway still remains: the use of facial recognition worsens racial discrimination in policing.

[16] See, for example, ARTICLE 19 (January 2021) Emotional Entanglement: China's emotion recognition market and its implications for human rights; ARTICLE 19 (2024), 'EU: AI Act passed in Parliament fails to ban harmful biometric technologies', 13 March; Leufer, D. (2021) 'Here's how to fix the EU's Artificial Intelligence Act', Access Now, 7 September.

[17] Interviews 1 and 8.

[18] Open letter of over 170 organisations.

[19] Electronic Frontier Foundation documented how 17 communities across the US have sought for facial recognition technology bans.

[20] EDRi's Reclaim Your Face campaign and engagement with MEPs led to the EU Parliament calling for bans of public facial recognition in the Artificial Intelligence Act. In this case, civil society advocacy resulted in a deal between the European Council Presidency and Parliament agreeing to a deal on the EU AI Act's text in December 2023, which included strong human rights protections. During final negotiations, big tech and start-up lobbying undermined human rights protections in the Act, resulting in a significantly weakened text that fails to draw effective red lines against biometric surveillance and predictive policing, among other glaring issues.

[21] The Brazilian Institute for Consumer Protection (Idec) and Nupef Institute called for a moratorium on facial recognition technologies in 2009. This led to the *Idec v ViaQuatro* case where a civil court in São Paulo held that ViaQuatro had violated data protection requirements and infringement of the right to privacy.

[22] The San Francisco Ordinance provides that it 'shall be unlawful for any Department to obtain, retain, access or use 1) any Face Recognition Technology; or 2) any information obtained from Face Recognition Technology on City-issued software or a City-issued product or device.'

[23] For example, Boston prohibits 'Boston or any Boston official' from '[e]nter[ing] into an agreement with any third party for the purpose of obtaining, retaining, possessing, accessing, or using, by or on behalf of Boston or any Boston official any face surveillance system; or [i]ssu[ing] any permit or enter[ing] into any other agreement that authorizes any third party, on behalf of Boston or any Boston official, to obtain, retain, possess, access, or use (i) any face surveillance system, or (ii) information derived from a face surveillance system.' See also Brookline, MA. The Minneapolis, MN, ban also includes a ban on '[e]nter[ing] into a contract with a third party that assists the third party in developing, improving, or expanding the capabilities of facial recognition technology or provides the third party with access to information that assists the third party in doing so.' The Portland, OR, ban notes that '[t]his prohibition applies to Face Recognition Technologies that are procured by any means with or without the exchange of monies or other consideration. For purposes of clarity, this means bureaus shall not purchase, lease or accept a donation or gift of Face Recognition Technologies. This prohibition applies to Face Recognition Technologies that are procured by any means with or without the exchange of monies or other consideration.'

[24] European Parliament (2023) 'MEPs ready to negotiate first-ever rules for safe and transparent AI', 14 June. The Committees of Internal Markets (IMCO) and Civil Liberties (LIBE) proposed a draft deal which would include 1) a total ban on the use of real time facial recognition and other biometric identification in public; 2) a ban on retrospective deployments of biometric identification systems with one exception for law enforcement, if they have judicial authorisation; 3) a ban on the sale, deployment or use of biometric categorisation systems which use sensitive characteristics like gender, race, ethnicity, etc.; 4) a ban on the sale, deployment or use of emotion recognition technologies by police, border authorities, employers or educational authorities; 5) a ban on the use of scraping tools for facial recognition databases. See European Parliament, Amendments adopted by the European Parliament on 14 June 2023.

[25] For example, Berkeley sets out a specific exemption 'for personal communication devices'. Other bans, such as in Brookline, make a similar exemption but specify that it covers 'or the sole purpose of user authentication' or 'for verification purposes or the sole purpose of user authentication' or 'for verification purposes'.

[26] For example, in Boston, this exemption is for '[u]sing social media or communications software or applications for communicating with the public, provided such use does not include the affirmative use of any face surveillance'.

[27] For instance, Brookline's and Boston's exemption for '[u]sing automated redaction software, provided such software does not have the capability of performing face surveillance'. Portland also provides an exemption for 'detecting faces for the sole purpose of redacting a recording for release or disclosure outside the City to protect the privacy of a subject depicted in the recording'.

[28] ACLU Louisiana (2022) ACLU of Louisiana Issues Statement After New Orleans City Council Reverses Surveillance Ban, Expands Use of Racists Technologies, 22 July. Berkeley and Cambridge require the 'receipt, access or use' to be logged in an annual surveillance report; see Lavoie, D. (2022) 'Virginia Lawmakers OK lifting ban on facial technology use', *AP News*, 10 March. Berkeley requires that 'all copies of the information are promptly destroyed upon discovery of the information, and the information is not used for any purpose'. However, it qualifies this requirement by noting that 'nothing in this Chapter shall limit the ability to use such information in connection with a criminal investigation'.

[29] In addition to the broad exemption contained within Berkeley's ban, Brookline includes an exemption for the use of 'evidence relating to the investigation of a specific crime that may have been generated from a face surveillance system'. Boston includes a similar exemption but requires that the 'evidence was not generated by or at the request of Boston or any Boston official'.

[30] Some bans contain exemptions for specific purposes such as '[c]omplying with the National Child Search Assistant Act'.

[31] While Article 5 prohibits the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement, it nonetheless subjects the prohibition to an exception, providing that: 'unless and in so far as such use is strictly necessary for one of the following objectives: (i) the targeted search for specific victims of ***abduction, trafficking in human beings or sexual exploitation of human beings, as well as searching for*** missing ***persons;*** (ii) the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or ***a genuine and present or genuine and foreseeable threat*** of a terrorist attack; (iii) the localisation ***or*** identification ***of a person suspected of having committed a criminal offence, for the purpose of conducting*** a criminal ***investigation, prosecution or executing a criminal penalty for offences referred to in Annex II*** and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least ***four*** years.'

[32] Expert meeting; the Electronic Frontier Foundation, on file.

[33] For example, in 2022, New Orleans reinstated the use of face recognition as an investigative tool citing rising violence in the city, although in a statement issued by the ACLU of Louisiana, it noted that the 'NOPD [New Orleans Police Department] and sponsors of the ordinance have admitted that there is absolutely no evidence that reinstating facial recognition will help reduce violence'. Similarly, Virginia amended the scope of its 2021 ban – initially covering local police and campus police from using face recognition unless explicitly authorised by law – to allow police agencies to use the technology under certain circumstances.

[34] For example, California introduced a three-year moratorium providing that a 'law enforcement agency or law enforcement officer shall not install, activate, or use any biometric surveillance system in connection with an officer camera or data collected by an officer camera'. However, the moratorium was not extended beyond 1 January 2023. At the national level, the Facial Recognition and Biometric Technology Moratorium Act is currently before Congress for a mapping of these bans. See Sheard, N. and Schwartz, A. (2022) 'The Movement to Ban Government Use of Face Recognition', Electronic Frontier Foundation, 5 May (contains links to each ban).

[35] In December 2021, Italy's Parliament also put in place a moratorium on public authorities and private entities' use of video surveillance systems that use facial recognition in public spaces, or places accessible to the public. The moratorium, initially proposed until December 2023, can be extended further until Italy passes a comprehensive law regulating facial recognition. See, for example, Carrer, L. (2021) 'The facial recognition moratorium passed in Italy reminds us why we need to call for a ban' (in Italian), Hermes Center for Transparency and Digital Human Rights, 2 December; and EDRi (2021) 'Italy introduces a moratorium on video surveillance systems that use facial recognition', 15 December.

[36] For example, the moratorium adopted by the Italian Parliament includes significant carve outs for judicial authorities and public prosecutors. In addition, it allows for police use of facial recognition technologies subject to a case-by-case approval by the Italian data protection authority (DPA).

[37] See, for example, Google's AI Principles, which states that Google 'will not design or deploy AI' in several application areas, including technologies that 'cause or are likely to cause overall harm', 'weapons or other technologies whose principal purpose or implementation is to cause or directly facilitate injury to people', technologies that 'gather or use information for surveillance violating internationally accepted norms', and, more generally, technologies whose purpose 'contravenes widely accepted principles of international law and human rights'. Since adopting the Principles in 2018, Google has published annual updates on its AI Principles. While these updates provide details on the processes developed within the company to implement the Principles and some substantive examples of how Google applies them, it has not yet published examples or confirmation of it, deciding against designing or deploying any particular technology after these Principles were put in place. See, for example, Rushe, D. (2018) 'Activists call for Salesforce boycott over US border patrol contract', *Guardian*, 20 August.

[38] For example, in 2018, employees at a number of major tech companies mobilised to call on their employers to either stop selling specific products to US government departments or not to bid for particular contracts. See, for example, An Open Letter to Microsoft: Don't Bid on the US Military's Project Jedi: Signed by employees of Microsoft, 13 October 2018; Wakabayashi, D. and Shane, S. (2018) 'Google Will Not Renew Pentagon Contract That Upset Employees', *New York Times*, 1 June; Fernandez, P. and Pendergrass, T. (2021) The Movement to End Police Violence One Year after George Floyd's Murder, 25 May; Amazon (2020), 'We are implementing a one-year moratorium on police use of Rekognition', 10 June; Jansen Reventlow, N. (2020) 'How Amazon's Moratorium on Facial Recognition Tech is Different from IBM's and Microsoft's', Slate Future Tense, 11 June.

[39] California Assembly Bill No. 331 (last amended 19 April 2023).

[40] A recent exception to this is Microsoft's responsible AI principles. They require companies wishing to use its facial recognition technologies to apply to Microsoft 'to prove they are matching Microsoft's AI ethics standards and that the features benefit the end user and society'. However, companies will be prohibited from using the technologies 'to infer emotional states and attributes such as gender or age'. San Francisco Board of Supervisors, Administrative Code – Acquisition of Surveillance Technology Ordinance No. 107-19 (21 May 2019).

[41] New Orleans backtracked on a facial recognition ban in 2022, and Virginia's state legislature enacted a bill that overturned a blanket ban to define use cases in which facial recognition technology can be used.

[42] See, for example, *R (Bridges) v CCSWP and SSHD*, [2019] EWHC 2341 (Admin) (4 September 2019), concerning the South Wales Police's deployment of facial recognition technologies in public spaces on two separate occasions. The Court of Appeal of England and Wales found that the interference with the right to privacy was not in accordance with the law as 'individual police officers' were given 'too much discretion' and '[i]t is not clear who can be placed on the watchlist nor is it clear that there are any criteria for determining where automated facial recognition (AFR) can be deployed'. It also found that the data protection impact assessment 'failed properly to assess the risks to the rights and freedoms of data subjects and failed to address the measures envisaged to address the risks arising from the deficiencies we have found [on the discretion accorded to individual police officers]'. Finally, the Court found that the South Wales Police 'have never sought to satisfy themselves, either directly or by way of independent verification, that the software program in this case does not have an unacceptable bias on grounds of race or sex'. It observed that, '[w]e would hope that, as AFR is a novel and controversial technology, all police forces that intend to use it in the future would wish to satisfy themselves that everything reasonable which could be done had been done in order to make sure that the software used does not have a racial or gender bias'. However, the Court underscored the limitations of judicial action by emphasising that it was only looking at the facts of the case presented to it and not any future hypothetical harms.

[43] In Brazil, a complaint brought by the Instituto Brasileiro de Defesa do Consumidor resulted in a company, ViaQuatro, being fined and the suspension of the use of its facial recognition technology which involves the capture of images, sound, and other personal data of everyday commuters through cameras or other devices without their prior consent on the São Paulo metro. See, for example, Global Freedom of Expression, Columbia University, 'The Case of São Paulo Subway Facial Recognition Cameras'; or Access Now (2021) 'Privacy win for 350,000 people in São Paulo: court blocks facial recognition cameras in metro', 12 May. As of December 2023, ViaQuatro lost its appeal and is now forced to pay a higher compensation fee which will be directed to the Fund for the Defense of Diffuse Rights (FDD). See commentary by the Brazilian Institute for Consumer Defense. It is noteworthy that the judgment states that should ViaQuatro resume such activities in the future, it must do so only after obtaining users' prior consent by providing clear and specific information about the collection and processing of data; Interview 2.

[44] In October 2023, the First-Tier Tribunal of the General Regulatory Chamber – Information Rights (the Tribunal) handed down its decision in Clearview AI Inc v The Information Commissioner [2023] UKFTT 819, overturning the fine. As of November 2023, the ICO is seeking permission to appeal the judgment of the First Tier Tribunal on the Clearview AI Inc case and is now awaiting the Tribunal's decision.

[45] Interview 2, on file.

[46] Interviews 2, 4, on file.

[47] Interview 2, on file.

[48] Interviews 1, 2, 3, 6, 8, on file.

[49] Interview 3, on file.

[50] Interview 3, on file.

[51] Expert meeting, on file.

[52] Expert meeting, on file. See also Investor Alliance for Human Rights, Investor Statement in Support of Digital Rights Regulations; and Morozov, E. (2013) *To Save Everything, Click Here: The Folly of Technological Solutionism*, New York, NY: PublicAffairs.

[53] For example, this was the case of the São Paolo metro system that used facial recognition technology to monitor people's reaction to advertising.

[54] Green, B. (2019) *The Smart Enough City: Putting Technology in its place to Reclaim our Urban Future*, Cambridge, MA: MIT Press; Prasad, M. and Marda, V. (2019) 'Interrogating Smartness: A case study on the caste and gender blindspots of the Smart Sanitation Project in Pune, India' Global Information Society Watch.

[55] Interview 1, on file.

[56] Expert meeting. One participant noted that '[i]n order to use such technology just to find, let's say one missing child … you need to scan whole people.'

[57] EDRi (2022) Civil society calls for the EU AI act to better protect people on the move, 6 December.

[58] The most prominent are those applications which purport to find missing children and promote healthcare access. During briefings with EU parliamentarians on the EU Artificial Intelligence Act, this was initially a common refrain from MEPs while also discussing emotion recognition.

[59] Many reports have now been issued highlighting baseline human rights concerns with new and emerging digital technologies generally and when deployed in specific use cases. For some of the initial reports, see ARTICLE 19 and Privacy International, *Privacy and Freedom of Expression in the Age of Artificial Intelligence*; Latonero, M. (2018) *Governing Artificial Intelligence: Upholding Human Rights & Dignity*, Data & Society; McGregor, L., Ng, V. and Shaheed, A. (2018) *The Universal Declaration of Human Rights at 70: Putting Human Rights at the Heart of the Design, Development, and Deployment of Artificial Intelligence*; ARTICLE 19 (2021) *An EU Artificial Intelligence Act for Fundamental Rights A Civil Society Statement*, 30 November; or ACLU of Massachusetts (2023) 'Press Pause on Surveillance'.

[60] See, for example, Amnesty International, Ban the Scan New York City and Ban the Scan Hyderabad.

[61] See, for example, ARTICLE 19, When bodies become data.

[62] See, for example, SHARE Foundation (2019) 'Huawei knows everything about cameras in Belgrade – and they are glad to share!', 29 March.

[63] For example, Google has not yet provided case studies on how its commitment not to 'design or deploy' harmful AI applies to concrete cases, including biometric technologies (see earlier).

[64] ARTICLE 19 (2024), 'EU: AI Act passed in Parliament fails to ban harmful biometric technologies', 13 March; EDRi and AI coalition partners (2024), 'EU's AI Act fails to set gold standard for human rights', 3 April.

[65] For example, in recognition that the risks presented by other forms of biometric technologies, such as voice and emotion recognition, are similar or even more severe than facial recognition technologies, some civil society actors have recently moved away from a sole focus on facial recognition technology to biometric technologies more broadly. See Berkeley.

[66] ARTICLE 19 (2021) 'EU: Risky biometric technology projects must be transparent from the start', 16 December; EDRi (2022) 'Regulating Migration Tech: How the EU's AI Act can better protect people on the move', 9 May.

[67] For example, securing a ban in San Francisco (see earlier) contributed to the adoption of further city bans in the US, a state moratorium, and commitments by some tech companies to stop selling facial recognition technologies to law enforcement agencies within the US until regulation is in place.

[68] For example, some interviewees noted that while there may be a need for wide bans across use cases, the focus of their particular initiatives were tied to their organisational mandates or the political context in which they operate. See, for example, ARTICLE 19 (2021) *Emotional Entanglement: China's emotion recognition market and its implications for human rights*; Stark, L. and Hutson, J. (2022) 'Physiognomic Artificial Intelligence', *Fordham Intellectual Property, Media & Entertainment Law Journal*, 32: 922–978. Similarly, others noted that even though they had heard of other problematic use cases, they were too far removed from the focus of their everyday work and therefore could not extend their call for AI red lines to other contexts. See earlier for some of these initiatives and the red lines they have resulted in.

[69] For instance, in the context of the EU Artificial Intelligence Act, remote real-time biometric mass surveillance was initially listed under unacceptable uses; however, this formulation left out *post facto* biometric surveillance. This was instinctively a problematic oversight from the vantage point of digital rights activists; however, it took significant work from civil society to underscore the importance of having both post and real time mass surveillance banned.

[70] This has been the case of city or state prohibitions or moratoriums, the draft EU Artificial Intelligence Act, and litigation or enforcement orders issued by regulatory bodies.

[71] A similar effort occurred in Europe, with Privacy International coordinating efforts with multiple NGOs.

[72] See, for example, UN Human Rights Council, *Surveillance and human rights, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, 28 May 2019, A/HRC/41/35. Similarly, the fixed nature of the draft EU Artificial Intelligence Act, as currently framed, not only creates risks for the design, development, and use of AI-enabled technologies within the EU, but also outside of it. This is because it is the first explicit regional framework regulating AI-enabled technologies and may therefore be used as a model elsewhere, leading to what has been referred to as the 'Brussels effect'. Accordingly, the nature of bans within the EU Artificial Intelligence Act carries direct consequences for the pursuit of AI red lines elsewhere. See Birhane and Raji, 'Chat GPT'; McQuillan, We come to bury ChatGPT'; or Weill, 'You Are Not a Parrot'; Harwell, Tiku and Oremus, 'Stumbling with their words'; Forbes, 'Microsoft confirms'; or Russell and Black, He's played chess.

[73] For example, following the ban of facial recognition technologies in San Francisco, other US cities adopted moratoriums and bans, although as discussed earlier, each is slightly different in nature. See, for example, Ryan-Mosley, T. (2023), 'The movement to limit face recognition tech might finally get a win', *MIT Technology Review*, 20 July.

[74] See, for example, the multiple legal actions brought against Clearview in different US states; Sheard and Schwartz, 'The Movement to Ban'.