

WITNESS Submission to the Joint Committee on Human Rights

This submission aims to provide input to the U.K. Parliament's Joint Committee on Human Rights [call for evidence](#) on Human Rights and the Regulation of Artificial Intelligence. The current call issued by the U.K. Parliament invites submissions on questions referring to (a) Human rights issues, (b) Existing legal and regulatory framework, as well as (c) Possible changes to legal and regulatory framework.

Executive Summary

WITNESS is submitting, therefore, comments to all three overarching points based on its years of experience in addressing the challenges of synthetic media and generative AI, specifically centering the perspectives of defenders, frontline journalists, and marginalized communities. WITNESS seeks to create a virtuous cycle where AI standards and legislation are robust, protect and preserve the needs and demands from frontline communities and journalists, and legislation embeds rights-respecting guardrails, ensuring that digital trust infrastructure protects those who are most at-risk and, by extension, benefits all users.

Current UK law is not sufficient to address these challenges. We recommend future legislation that treats provenance and authenticity as digital public infrastructure, embeds privacy, equity and accessibility into standards and tools, and ensures detection systems are evaluated for real-world effectiveness through socio-technical benchmarks such as TRIED. Liability must attach across the AI lifecycle, with timely redress for victims of harm, and regulation must be dynamic and internationally aligned, enabling the UK to play a leadership role in safeguarding human rights as AI evolves.

About WITNESS

WITNESS brings over 30 years of pioneering expertise and experience at the intersection of human rights, video, technology, and citizen journalism. Our guidance on how to document, verify and prove real, and advocate in an audiovisual world is used by millions globally and sets the industry standards. Throughout multiple technical shifts and transformations, we've successfully worked with our core stakeholders to maintain their credibility and impact as both frontline voices and key systems actors. This is now more important than ever as AI radically shifts our core understanding of how audiovisual content is created, verified, and interpreted.

For the past eight years, we have led proactive efforts to address the challenges of synthetic media and generative AI, specifically centering the perspectives of defenders, frontline journalists, and marginalized communities. This experience and foresight, long before deepfakes entered mainstream awareness, uniquely positions WITNESS to bridge between

audiovisual truth-telling and the current AI transformation influencing our information environment:

- **Pioneering Anticipatory Response:** Initiated “[Prepare. Don’t Panic](#)” in 2018, the first broad civil society effort to anticipate deepfake threats, ensuring global civil society inclusion in critical policy and influencing global discussion on how to understand and respond to AI-generated threats.
- **Global Expertise and Reach:** Built a globally distributed team across 5 regions that partners with community media, journalists, and human rights defenders in 130+ countries, grounding our work in diverse lived experiences and responsive to real-world needs across different information environments.
- **Community-Led Verification:** Launched the “Fortifying Community Truth” network, beginning in West Africa, and developed the “Community-Based Approach to Verification Guide”, grounded in deep expertise in reinforcing factual information and challenging mis/disinformation through locally-appropriate methods.
- **Deepfakes Rapid Response Force:** Established the first [global mechanism](#) for forensic analysis of suspected deepfakes to support frontline journalists and fact-checkers in real-time, combining it with leading-edge training on detecting deceptive AI.
- **AI Standards Development & Policy Influence:** Shaped global [technical standards](#) for media and [content authenticity](#) as well as ethical norms for how media, civil society, and other stakeholders should transparently disclose and responsibly use generative AI, strengthening this work with our [high-level legislative engagement](#) and public discourse leadership via [high-profile events](#) and media.

WITNESS Submission to the Joint Committee on Human Rights.....	1
Executive Summary.....	1
About WITNESS.....	1
Human Rights Issues.....	4
1. How can Artificial Intelligence (AI) affect individual human rights for good or ill, in particular in the areas of: (a) privacy and data usage; (b) discrimination and bias; and (c) effective remedies for violations of human rights?.....	4
Privacy and data usage.....	4
Discrimination and bias.....	4
Effective remedies for human rights violations.....	5
Existing legal and regulatory framework.....	6
2. To what extent does the UK’s existing legal framework provide sufficient protections for human rights in relation to AI?.....	6
3. To what extent is the Government’s policy approach to deploying AI, expressed in its “AI Opportunities Action Plan”, sufficiently robust in respect of safeguarding human rights?.....	6
Transparency and provenance.....	8
Evaluation of detection tools.....	8

Participation and accountability.....	8
Possible changes to legal and regulatory framework.....	8
4. What would be needed in any future UK legislation to protect human rights?.....	8
• To what extent should the same human rights standards apply to private actors as public bodies when they use AI?.....	8
• To what extent might different kinds of AI technology require different regulatory approaches?.....	8
Provenance and authenticity as digital public infrastructure.....	8
Evaluation and accountability for detection systems.....	9
Participation and oversight.....	9
Risk-based approach.....	10
5. Who should be held accountable for breaches of human rights resulting from uses of AI, and on what basis?.....	10
• Where in the process of developing, deploying and using AI technologies should liability arise?.....	10
• What additional measures, if any, are needed to ensure that individuals have sufficient redress where they have suffered harm because of the use of AI?.....	10
Where liability should arise	
Liability should not be confined to end-use. It should attach at three stages:.....	11
Ensuring sufficient redress	
At present, individuals harmed by AI-generated or manipulated content often lack effective remedies. To address this, legislation should:.....	11
6. How might regulation match the pace of AI technology development, such as the emergence of agentic AI, to ensure that human rights are preserved as technology continues to develop?.....	11
Dynamic standards and conformance.....	12
Iterative evaluation frameworks.....	12
Anticipatory governance for emerging risks.....	12
7. How could regulation take account of the international nature of AI? How could it address the potential consequences for human rights in the UK of the malign use of AI by regimes in other countries?.....	12
Alignment with international standards and legislation.....	12
Global accessibility of trust infrastructure.....	13
Cross-border accountability and redress.....	13
A leadership role for the UK.....	13
8. How much difference will the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law make to the protection of human rights in the UK?.....	14
9. What lessons can be drawn from regulation of the impact of AI on human rights in other jurisdictions, such as the European Union?.....	14
European Union: embedding provenance and transparency.....	14
California: clarity on provenance standards.....	14
Global standards bodies: embedding human rights in technical infrastructure....	14

Human Rights Issues

1. How can Artificial Intelligence (AI) affect individual human rights for good or ill, in particular in the areas of: (a) privacy and data usage; (b) discrimination and bias; and (c) effective remedies for violations of human rights?

Privacy and data usage

- Artificial intelligence is rapidly transforming the information environment, adding layers of both capability and complexity. On one hand, AI systems can process vast amounts of data, generate insights, and create content at unprecedented speed and scale. On the other, this same power introduces risks of deception—ranging from subtle misinformation shaped by biased training data to deliberately crafted deepfakes and synthetic texts designed to manipulate perception. This can take the form of highly personalized misinformation streams or deepfakes that exploit individuals’ likenesses, eroding trust in what we see and hear. In this environment, provenance and authenticity infrastructure becomes critical—not only to signal when AI has been used, but also to provide verifiable proof that content is real and unaltered.
- Provenance is the source and history of the media we consume online. If captured in an accessible, rights-respecting and verifiable way, it can help us to trace the origin, chain of custody, and authenticity of media. It can inform about wholly AI-generated material, communication that mixes AI and human inputs, as well as the provenance of entirely ‘real’ content. WITNESS has been at the forefront of shaping how provenance technologies are developed, regulated and deployed, particularly with global consortia and international standards bodies
- Provenance and authenticity infrastructures can help to safeguard trust in content, but if poorly designed they risk exposing sensitive personal data. In the Coalition for Content Provenance and Authenticity (C2PA), where WITNESS co-chairs the Threats and Harms Working Group, we have consistently argued for a “how, not who” approach. Provenance should disclose how content was created or altered, not who created it, to prevent surveillance and protect anonymity. We have embedded privacy protections into C2PA specifications and argued for accessibility to open-source and small-scale implementations, to prevent provenance from becoming a privilege available only to well-resourced actors.

Discrimination and bias

- Under the topic of bias and discrimination, when amplified by AI-driven content and deepfakes, these aspects can present significant risks to democratic integrity. Algorithms trained on skewed or incomplete data can disproportionately target certain communities with misleading narratives, suppressing voter participation or reinforcing harmful stereotypes. At the same time, deepfakes—whether of political candidates, activists, or ordinary citizens—can be weaponized to spread falsehoods that exploit existing social divides, often preying on marginalized groups who already face systemic discrimination.

This combination not only distorts public perception but also undermines trust in the electoral process itself, making it harder for voters to access accurate, unbiased information on which to base their decisions.

- Deepfakes detection tools are vital for safeguarding trust in information, helping identify and flag manipulated content before it can mislead the public or undermine democratic, journalistic, and human rights efforts.
- Detection tools are often assessed on narrow technical benchmarks, which overlook how they fail in diverse linguistic, cultural and low-quality contexts. Through the [Deepfakes Rapid Response Force \(DRRF\)](#), WITNESS has analysed suspected synthetic media in elections and conflicts worldwide. These cases show how existing tools underperform in African languages, noisy audio or compressed video. To address this gap, we developed the [TRIED Benchmark](#), a socio-technical evaluation framework that measures not only accuracy but also usability, explainability and accessibility. TRIED is now being taken up in policy and standards discussions as a means of ensuring detection tools meet real-world needs.

Effective remedies for human rights violations

- Human rights defenders (HRDs) and journalists rely on credible evidence—witness accounts, photos, videos, and written reports—to expose abuses and seek justice. However, as generative AI tools become more sophisticated and accessible, governments and perpetrators of violations increasingly dismiss or discredit authentic documentation by claiming it could be fabricated by AI. This can undermine trust in legitimate evidence, erodes the credibility of advocates, increases the burden of proof on human rights defenders and weakens accountability mechanisms in courts, international institutions, and the public sphere. In fact, the very technologies designed to enhance communication and documentation can also risk compromising the protection of human rights by providing abusers with plausible deniability and making it harder for survivors and defenders to have their voices heard and believed.
- Victims of non-consensual synthetic sexual imagery, political disinformation or fabricated “evidence” often lack clear redress. WITNESS integrates frontline experiences into standards and policy. Evidence from DRRF escalations and regional training in Latin America, West Africa and the East Asia is being fed into further refining our AI Detection benchmark (TRIED) and into standards-setting at C2PA and the ITU’s AI and Multimedia Authenticity Standards initiative, to ensure that remedies and redress are informed by actual harms.
- *Through WITNESS’s extensive work on risk and harm assessments of provenance and authenticity standards, alongside advocacy for equitable and effective detection tools, we have seen firsthand the limitations of these technologies in responding to TFGBV. In many cases, whether the content is AI-generated or manipulated does not change the harm inflicted—damage to reputation, credibility, safety, and personal security is already done, with little recourse for those affected. This form of violence disproportionately targets community leaders and human rights defenders, who continue to advocate and prioritize addressing this threat despite systemic neglect. Yet, responses from*

governments and technology companies remain inadequate, particularly for marginalized individuals who lack the visibility or verification mechanisms that automated systems rely on. As a result, ineffective technological solutions fail to provide meaningful protection, leaving critical gaps in accountability and safety.

- On the note of Elections and Conflicts, the weaponization of generative AI in elections and conflicts is increasingly targeting women in public roles, with major concerns including: (a) The use of non-consensual deepfakes and AI-driven sexual and gender-based violence (SGBV) to silence, discredit, and intimidate female politicians, activists, community leaders, and journalists—ultimately undermining women's participation in democracy and public discourse; and (b) The long-term chilling effect on freedom of expression due to the widespread deployment of AI-generated non-consensual intimate imagery (NCII), deterring women from engaging in public life. Our advocacy in legislative processes, including the EU AI Act and California Assembly Bill 853, stresses that provenance and detection cannot themselves be definitive solutions. They must be accompanied by legal obligations for timely takedowns, transparency and accountability that give survivors and affected communities real avenues for remedy. There is a need for the UK government to be ahead of these trends and potential harms, and initiate things such as the UK's own Provenance Bill.

Existing legal and regulatory framework

2. To what extent does the UK's existing legal framework provide sufficient protections for human rights in relation to AI?

3. To what extent is the Government's policy approach to deploying AI, expressed in its "AI Opportunities Action Plan", sufficiently robust in respect of safeguarding human rights?

The UK's current legal framework provides partial protection but leaves significant gaps when applied to AI. Legislation such as the Data Protection Act 2018 and the Equality Act 2010 offer important safeguards on privacy and discrimination, but they are not tailored to the unique challenges of AI-mediated audiovisual content. Added to that, we believe the current framework does not yet provide sufficient or comprehensive protections for human rights in relation to AI. Future legislation will need to incorporate explicit safeguards on provenance, detection and equitable access, as well as mechanisms for accountability and redress.

While looking at the Government's "AI Opportunities Action Plan" it remains clear that the document is primarily oriented towards economic growth and innovation. While it recognises risks in a general sense, it does not set out a robust framework for safeguarding human rights. The idea that [regulation or embedding human rights](#) into the development of new products and technologies is opposed to innovation is a dangerous fallacy. Trust in new solutions can increase when supported by robust frameworks that take into account aspects such as privacy, freedom of expression, non-discrimination, content authenticity and provenance. Developing accountable, transparent and trustworthy solutions, with companies acting responsibly in

addressing the harm stack should remain the focus, together with the newer pushes for innovation.

When it comes to current legal framework's gaps, we would like to point out the following critical areas where the UK framework is insufficient:

- **Transparency:** AI regulations that emphasize transparency can play a crucial role in protecting users by requiring clear disclosure of when and how AI is involved in generating or shaping content. Such measures help people better understand the origins and reliability of the information they encounter, reducing the risks of deception, bias, or manipulation. By mandating accountability and explainability in AI systems, regulations and tools such as labels to AI-generated or manipulated content, we can safeguard individual rights and reinforce public trust in digital platforms and the broader information ecosystem.
- **Provenance and authenticity:** At present, there are no clear requirements for the use of provenance and authenticity infrastructure in the UK. By contrast, the European Union's AI Act (Article 50) establishes obligations for transparency and provenance in AI-generated content. In the United States, California legislation has already advanced rights-respecting provenance standards. Without equivalent provisions, the UK risks falling behind in protecting the integrity of audiovisual information and in aligning with international standards.
- **Evaluation of AI detection tools:** The UK has no mechanism to assess whether AI detection systems perform effectively across diverse contexts. WITNESS's TRIED Benchmark shows that most tools, when tested, perform poorly in real-world environments and especially in under-resourced communities. Without frameworks for socio-technical evaluation, individuals will be left with unreliable tools, limiting their ability to seek redress or protect their rights.
- **Access and equity:** The current framework does not address the risk of creating a two-tier information environment in which provenance and detection tools are accessible only to well-resourced actors. Our work with frontline journalists and fact-checkers demonstrates that without guarantees of accessibility and usability, protections remain uneven, undermining equality before the law.

With regards to the Action Plan, we trust it must provide sufficient assurance that AI deployment in the UK will uphold human rights. Currently it does not. Stronger provisions are needed to embed transparency and provenance, to evaluate detection systems against real-world use cases, and to guarantee inclusive participation in governance processes. Added to that, we would also like to highlight three areas of particular concern:

Transparency and provenance

- The Action Plan does not adequately address the role of provenance and authenticity infrastructures as essential safeguards for human rights and democratic trust. Without clear commitments in this area, the UK risks falling behind other jurisdictions such as the EU, where Article 50 of the AI Act mandates transparency measures for AI-generated content. WITNESS's experience in standards bodies such as C2PA and ITU AMAS shows that embedding privacy, equity and usability at the infrastructure level is critical to ensuring these systems work in practice.

Evaluation of detection tools

- The Action Plan does not account for the limits of current detection technologies or provide mechanisms to ensure that tools are effective and accessible. Our work through the TRIED Benchmark demonstrates that detection systems frequently underperform in real-world conditions, particularly in diverse linguistic and cultural contexts. Unless government policy recognises and responds to these limitations, public reliance on detection will be misplaced, leaving people without effective protection.

Participation and accountability

- The Action Plan does not embed the perspectives of those most at risk from AI misuse, including journalists, human rights defenders and marginalised communities. Through initiatives such as the Deepfakes Rapid Response Force and regional trainings in Africa, Latin America and Southeast Asia, WITNESS has seen how the voices of frontline actors provide essential insights for building rights-respecting systems. Government policy should mandate participatory processes and ensure that civil society expertise is included in standards, regulation and oversight.

Possible changes to legal and regulatory framework

4. What would be needed in any future UK legislation to protect human rights?

- To what extent should the same human rights standards apply to private actors as public bodies when they use AI?
- To what extent might different kinds of AI technology require different regulatory approaches?

Future UK legislation should embed clear, enforceable safeguards that protect human rights across the full lifecycle of AI systems. Three areas are particularly important.

Provenance and authenticity as digital public infrastructure

- Legislation should establish provenance and authenticity systems as a form of digital public infrastructure, designed around principles of privacy, accessibility and equity.

Additionally, through the widespread use of provenance and authenticity systems, we can fight against deceptive content through the promotion of a more resilient digital environment with more well informed users.

Provenance should focus on *how* content was created or altered rather than *who* created it, to protect anonymity and avoid surveillance. And, in order for that to be achievable, we need an updated legal framework that incorporates the need for such tools, while preserving privacy and safeguarding users from high-risk cases. Lastly, the inclusion of provisions around the need for certain systems to store *system provenance data* (technical information that helps verify authenticity) instead of *personal provenance data* (which could expose individuals) is a distinction that ensures that provenance systems do not compromise privacy or create risks of surveillance. Similar clarity is needed in the UK to prevent poorly designed provenance regimes from harming rights or creating inequitable access.

Evaluation and accountability for detection systems

- Legislation should acknowledge the relevance of detection tools for addressing mis/disinformation promoted through synthetic media and require that AI detection tools are assessed not only on technical accuracy but also on usability, accessibility and contextual reliability.

On the point of detection tools, WITNESS developed the TRIED Benchmark, which goes beyond performance metrics to evaluate tools in real-world conditions, including diverse languages, compressed media and frontline workflows. TRIED has been cited in policy advocacy in the EU and the US as a model for how to embed socio-technical evaluation into law and standards. A UK framework could adopt this approach to ensure that detection tools are genuinely usable in high-stakes contexts such as elections and conflicts.

Participation and oversight

- Future regulation must embed civil society and frontline perspectives into governance processes, recognising that standards are not neutral.

In the EU AI Act, WITNESS [has called](#) for downstream transparency obligations and fundamental rights impact assessments for high-risk AI systems, ensuring that the perspectives of those most at risk are considered when AI systems are deployed.

In the C2PA, [we are embedding harm assessments into the conformance programme](#), so that tools which cause or enable harm can lose certification. This creates an enforceable accountability mechanism that ensures standards evolve alongside real-world risks.

Risk-based approach

- A risk-based approach to AI helps protect human rights by focusing oversight and safeguards on the highest-risk applications, ensuring that potential harms like discrimination, surveillance, or manipulation are identified and mitigated before they can impact people's lives.

Different categories of AI technology will require different regulatory approaches. High-risk systems, such as those used in elections, biometric surveillance or provenance and authenticity infrastructure, warrant stricter safeguards, transparency obligations and independent oversight. Lower-risk applications may require lighter regimes, provided that they still comply with baseline human rights standards.

5. Who should be held accountable for breaches of human rights resulting from uses of AI, and on what basis?

- Where in the process of developing, deploying and using AI technologies should liability arise?
- What additional measures, if any, are needed to ensure that individuals have sufficient redress where they have suffered harm because of the use of AI?

Accountability and liability must be distributed across the full AI pipeline. Human rights breaches linked to AI rarely result from a single actor's decision, but from design choices, deployment practices and governance failures that combine to produce harm.

- Developers of models should be held accountable where design decisions foreseeably create risks to privacy, equity or accessibility. For example, provenance systems designed without privacy-by-design safeguards could expose identities or enable surveillance. In the Coalition for Content Provenance and Authenticity (C2PA), WITNESS has argued for privacy-preserving specifications that limit the capture of sensitive data. Without these safeguards, developers would bear responsibility for harms arising from misuse.
- Platforms that adopt detection systems should be obliged to integrate socio-technical evaluation as a component in their usage.
- Manufacturers and distributors of devices embedding provenance and authenticity features should be accountable for ensuring that users can opt in or out of sharing data, and that sensitive information is not disclosed by default. WITNESS' recommendations to amend California Assembly Bill 853 stressed the need to distinguish between *system provenance data* and *personal provenance data*. Manufacturers could expose users to surveillance or identity-based targeting without sufficient provisions in the UK.

- Standards bodies and regulators also carry responsibility. If certification and conformance programmes fail to include harm assessments, harmful systems will be legitimised. WITNESS has worked to embed such harm assessments into the C2PA conformance programme, ensuring that tools which endanger privacy or equity can lose certification. Regulators should be accountable for mandating this kind of oversight.

Where liability should arise

Liability should not be confined to end-use. It should attach at three stages:

1. **Design:** when foreseeable harms are ignored in development, such as provenance systems that collect unnecessary personal data.
2. **Deployment:** when AI is rolled out without adequate safeguards, evaluation or transparency, as seen in elections where faulty detection tools caused mislabelling.
3. **Distribution:** when platforms fail to act on evidence of harm or misuse, leaving victims without remedies.

Ensuring sufficient redress

At present, individuals harmed by AI-generated or manipulated content often lack effective remedies. To address this, legislation should:

- Mandate timely takedown and reporting mechanisms for AI-generated non-consensual intimate imagery and deceptive content.
- Require companies to provide clear explanations of provenance and detection outputs, so that users can understand and contest decisions. WITNESS' [TRIED Benchmark](#) highlights explainability as essential to fairness.
- Establish accessible routes to independent review and redress, ensuring that remedies are not left solely to platform discretion.

6. How might regulation match the pace of AI technology development, such as the emergence of agentic AI, to ensure that human rights are preserved as technology continues to develop?

AI systems are developing at a pace that outstrips traditional regulatory cycles. Static frameworks will struggle to keep up, especially with the rapid emergence of multimodal and agentic AI. The UK should therefore adopt flexible and anticipatory mechanisms that can evolve alongside technology, while grounding them in human rights principles.

Three approaches are particularly important:

Dynamic standards and conformance

- Standards and certification schemes must be living frameworks, updated in response to evidence of harm. WITNESS has argued within the Coalition for Content Provenance and Authenticity (C2PA) that conformance programmes should embed regular harm assessments. If a tool or implementation is shown to cause harm, it should lose certification. This creates a mechanism for standards to adapt in real time, rather than relying on slow legislative reform.

Iterative evaluation frameworks

- The TRIED Benchmark, developed by WITNESS, provides a socio-technical method for evaluating AI detection tools in real-world contexts. TRIED is being updated to reflect new modalities, such as audio deepfakes and multimodal synthetic media, through cases escalated to the Deepfakes Rapid Response Force. Embedding such iterative evaluation frameworks into regulation would allow the UK to ensure that detection and provenance systems remain fit for purpose as technology evolves.

Anticipatory governance for emerging risks

- Our monitoring on contextual AI and wearables highlights the need for anticipatory regulation that considers how AI will affect privacy and autonomy in dynamic, real-time environments. For example, AI-enabled wearables will continuously collect and process biometric data, requiring adaptive consent mechanisms and stronger safeguards against manipulation and surveillance. Similar anticipatory frameworks are needed to address the risks of agentic AI, where autonomous systems can act with minimal human oversight.

By integrating dynamic standards, iterative evaluation, and anticipatory governance, the UK can create a regulatory approach that is resilient to technological change while firmly rooted in the protection of human rights.

7. How could regulation take account of the international nature of AI? How could it address the potential consequences for human rights in the UK of the malign use of AI by regimes in other countries?

Alignment with international standards and legislation

- The UK should ensure compatibility with frameworks such as the EU AI Act (particularly Article 50 on transparency and provenance), the Council of Europe Framework Convention on AI and Human Rights, and international standards-setting efforts at ITU and ISO. WITNESS has contributed directly to these processes through its leadership

role in the Coalition for Content Provenance and Authenticity (C2PA) and in the AI and Multimedia Authenticity Standards group (AMAS). Alignment will help ensure interoperability, avoid regulatory gaps, and provide clarity for companies operating across borders.

We also suggest such alignment also addresses points such as Technology Facilitated Gender based violence (TFGBV) and the emerging use of nudifying apps; challenges emerging from the use of Synthetic media, deepfakes, and multimodal generative AI; the impact of emerging technologies in elections integrity and conflicts; and, last but not least, transparency across the board of AI systems.

Global accessibility of trust infrastructure

- Provenance and detection systems must not become tools available only in well-resourced markets. WITNESS has warned of the risks of a two-tiered information environment, where only certain regions or communities have access to authenticity infrastructure. If the UK adopts provenance or detection requirements, it should advocate internationally for open and equitable access, ensuring that frontline journalists and human rights defenders worldwide benefit from the same protections.

Cross-border accountability and redress

- Malign uses of AI abroad, including election manipulation and state-sponsored disinformation, have direct impacts on UK information ecosystems. WITNESS's Deepfakes Rapid Response Force has documented how AI-manipulated media originating in one country can rapidly spread across borders. UK regulation should establish mechanisms for international cooperation, including shared datasets, common benchmarks such as TRIED, and coordination on redress processes.

A leadership role for the UK

- By embedding privacy, anonymity and equity safeguards into any provenance or authenticity legislation, the UK could set a higher bar than current international practice. WITNESS's advocacy in California and the EU shows that early legislative clarity on the distinction between *system provenance data* and *personal provenance data* is critical to protecting rights. A UK provenance framework that takes this approach would not only protect domestic users but also set an influential model for others.

By embedding international alignment, accessibility and cooperation into its framework, and by leading with a rights-based model of provenance and detection, the UK can strengthen global protections while ensuring its own citizens are safeguarded.

8. How much difference will the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law make to the protection of human rights in the UK?

The Council of Europe Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law represents the first binding international treaty on AI. It provides a valuable baseline of principles, particularly in emphasising the need for AI systems to uphold human rights and democratic values. The Convention can strengthen protections in the UK, but only if it is treated as a floor rather than a ceiling. The UK should see it as a starting point, specifically due to the human-rights based approach taken on by the resolution, and move further by embedding specific safeguards on provenance, detection and accountability into its own legal framework.

9. What lessons can be drawn from regulation of the impact of AI on human rights in other jurisdictions, such as the European Union?

There are valuable lessons for the UK from the European Union, the United States, and global standards-setting bodies where WITNESS has been directly engaged.

European Union: embedding provenance and transparency

- The EU AI Act, in particular Article 50, introduces obligations on transparency and provenance for AI-generated content. This creates a foundation for downstream accountability and aligns technical standards with human rights objectives. WITNESS has contributed to shaping this debate, including through positions on the EU General Purpose AI Code of Practice and advocacy around Article 50 implementation. The UK can learn from this by ensuring that provenance obligations are framed as rights-preserving infrastructure rather than narrow compliance exercises.

California: clarity on provenance standards

- In California, Assembly Bill 853 builds on the state's AI Transparency Act by introducing provenance requirements. WITNESS's interventions highlighted the importance of distinguishing between *system provenance data* and *personal provenance data*, to prevent privacy harms and avoid creating surveillance risks. The UK can adopt this clarity to ensure that provenance frameworks protect users without exposing identities.

Global standards bodies: embedding human rights in technical infrastructure

- WITNESS's leadership in the Coalition for Content Provenance and Authenticity (C2PA) and the ITU's AI and Multimedia Authenticity Standards initiative (AMAS) shows the importance of embedding civil society perspectives early in technical standard-setting. Lessons from these forums demonstrate that standards are not neutral: they must be designed with privacy, accessibility and equity in mind. UK regulators should engage

directly with standards bodies to ensure domestic legislation is interoperable and globally relevant.