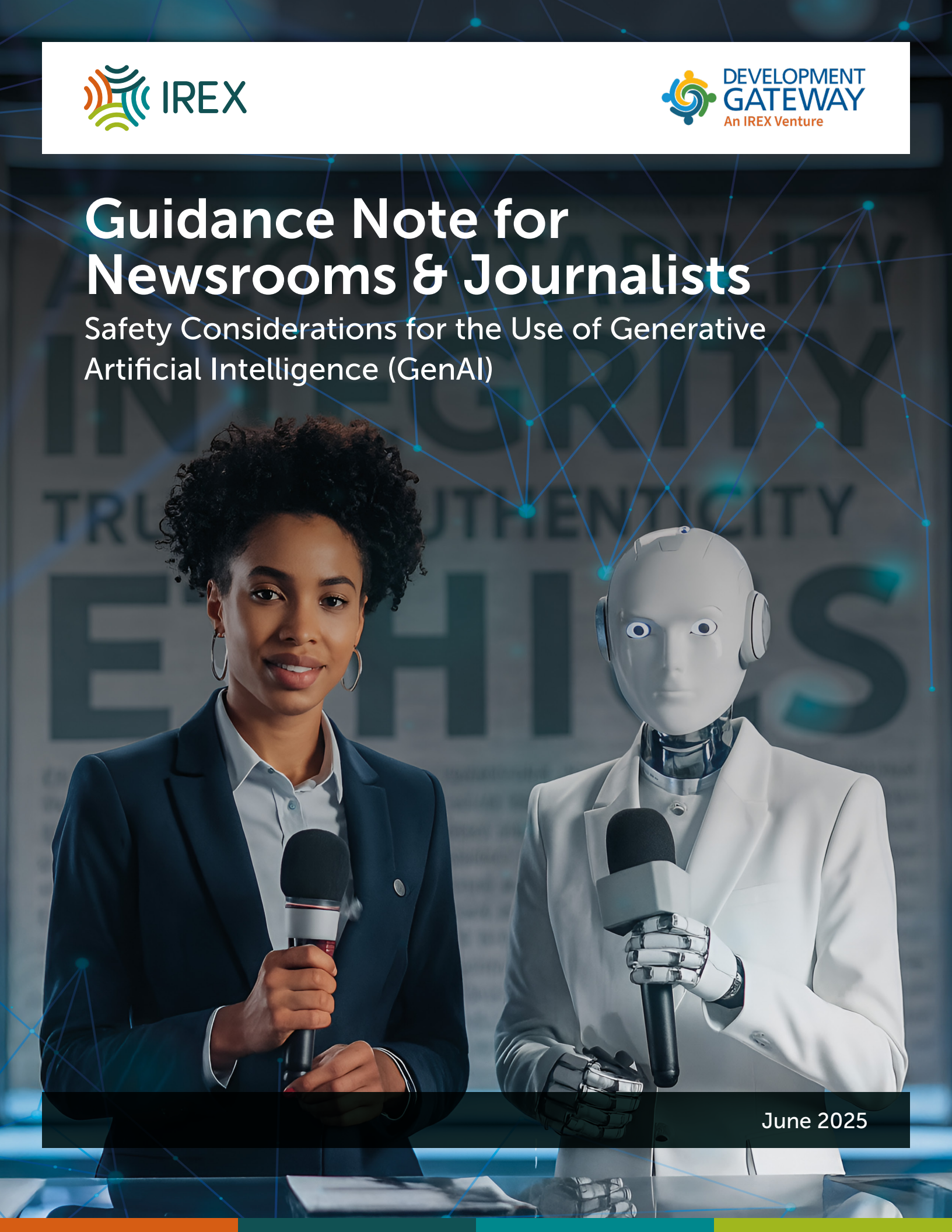
IREX

DEVELOPMENT GATEWAY
An IREX Venture

Guidance Note for Newsrooms & Journalists

Safety Considerations for the Use of Generative Artificial Intelligence (GenAI)



June 2025

Contents

This guidance note is designed to help journalists and newsrooms integrate generative artificial intelligence (GenAI) tools into their work safely, responsibly, and effectively. It outlines key applications and related risks, and provides clear, practical steps on how to use GenAI in a way that enhances, rather than undermines, independent journalism.

In developing this guidance note, IREX and Development Gateway do not seek to advocate for or against its use, but rather to provide practical guidance to help journalists and newsrooms safely adapt to this emerging reality, in line with our commitment to responsible AI (RAI).[1]

The development of GenAI has advanced at a staggering pace, moving from experimental technologies to mainstream tools in just a few years, and transforming how content is created, produced, and consumed. While GenAI remains a relatively new field, the landscape of associated risks and opportunities continues to evolve rapidly.

It is essential that media outlets and newsrooms stay ahead of these developments and engage with GenAI in a safe, informed, and responsible manner. By actively identifying and mitigating potential risks, journalists and media organizations will be better positioned to harness the significant potential of GenAI to enhance their work and uphold ethical standards.

This guidance note is intended to be a living document. It will be updated regularly to reflect the fast-changing nature of GenAI, helping to ensure that the advice remains relevant, practical, and aligned with emerging trends and challenges.

KEY APPLICATIONS: HOW GENAI IS CHANGING THE NEWS MEDIA

- Research Support
- Content Creation
- Dynamic Visuals
- Audience Engagement
- Workflow Efficiency
- Innovation and Experimentation
- Accessibility
- Business Development

See full descriptions in Annex 2

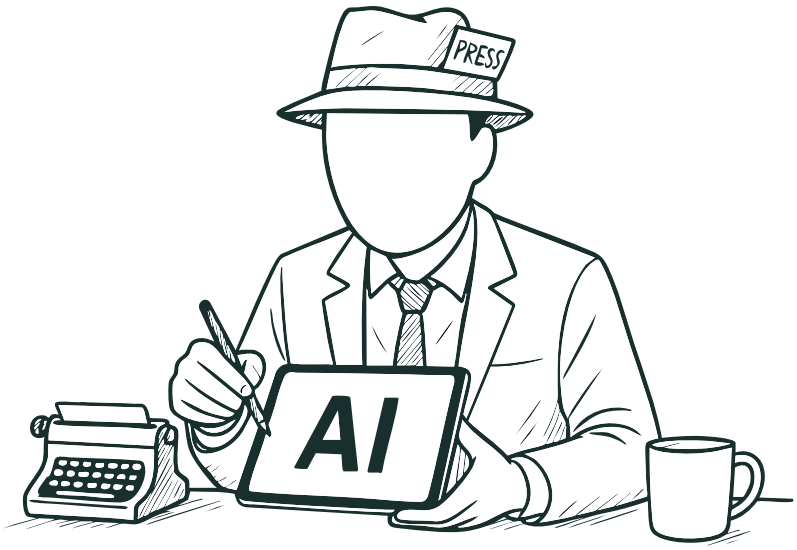
1. RAI refers to the development and use of AI in a manner that prioritizes safety, fairness, accountability, transparency, and respect for human rights. It emphasizes minimizing risks and harms while maximizing the positive impact of AI technologies.

Navigating the Risks of GenAI in Journalism

Integrating GenAI into journalism demands thoughtful oversight to ensure that transparency, accuracy, and editorial integrity remain at the core of the profession. As we explore the potential of these tools, it is equally important to assess and manage the risks they introduce. As GenAI's capabilities expand, so too do the stakes for how the media choose to adopt and apply them.

For journalists and media organizations, the risks associated with GenAI broadly fall into three interrelated categories: **editorial, legal,** and **digital security.** Each presents distinct challenges—but none exist in isolation.

These categories are deeply interconnected, and as GenAI technologies continue to evolve, a holistic approach to risk management becomes essential. Addressing these challenges collectively, while also considering their wider societal impacts, is critical to protecting the integrity, independence, and public trust in journalism in a future shaped by AI.



Editorial Risks

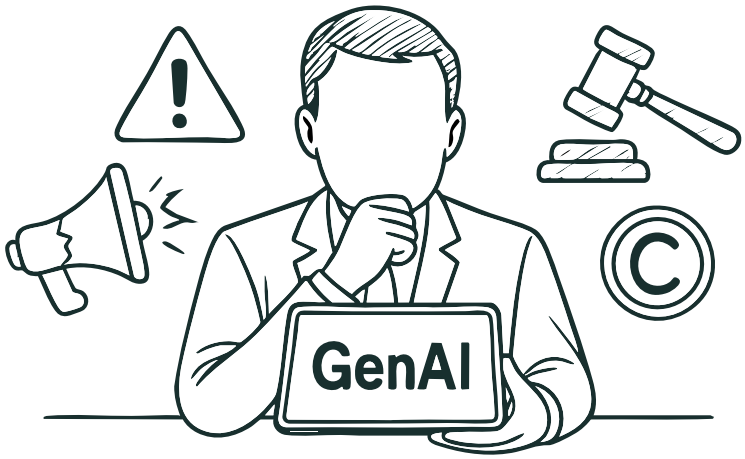
The use of GenAI in journalism raises serious editorial risks that go to the heart of trust and journalistic integrity. The crux of the issue is that GenAI outputs are produced by synthesizing vast, often undisclosed data sets that are largely trained on human-generated data and, as a result, shaped by human biases. Users typically lack insight into the data sets the model was trained on or indeed the methods it employs to prioritize information. When GenAI tools are used to support research or generate text summaries, it can be nearly impossible to trace the origin of the information. This also raises serious concerns about the potential for manipulation, including through techniques like data poisoning or prompt injection (see Annex 1, Glossary). Additionally, GenAI is prone to producing "hallucinations," that is, content that is factually incorrect, misleading, or entirely fabricated. In an era where elections, conflicts, and public health crises depend on citizens having access to accurate and trustworthy information, the unregulated use of GenAI in journalism therefore presents significant risks. The risk of plagiarism is another significant threat, as AI may unintentionally replicate proprietary content without attribution. Compounding this is the uncertainty around the copyright status of AI-generated output itself (see the "Legal Risks" section below). Journalists need to be aware of what rights they may be giving up when using GenAI tools, especially when it comes to copyright. They also need to understand the differences between using free tools and paid ones, as the terms and conditions can vary significantly, with implications regarding data privacy and copyright, and so it is important to read the small print. Moreover, since public GenAI systems draw from shared data sets, there is a growing risk of homogenization. Content across different outlets can demonstrate uniformity in voice, running the risk of reinforcing dominant narratives, flattening originality, diluting brand identity, and dampening innovation and social progress. This standardization, coupled with the potential for AI to reflect or amplify biases, can erode public trust, especially if outputs unintentionally reproduce discriminatory or extremist viewpoints. It also presents a complex challenge to the core journalistic value of authenticity.

As access to powerful generative tools becomes widespread, GenAI makes disinformation cheap to produce and disseminate, and increases its sophistication. The reality is that journalists and media outlets must compete for audiences' attention and revenue in this drastically disrupted information space. Moreover, AI-generated content can be used to impersonate real journalists or media brands, undermining legitimate reporting and eroding

public trust. AI-generated images and deepfakes are also being used to manipulate narratives during high-stakes critical moments,[2] creating confusion that makes it difficult for journalists and information consumers to discern fact from fiction. Journalism's credibility depends on transparency and trust. The catch-22 is that research has shown that labeling content as AI-generated can actually undermine trust, with audiences tending to perceive such material as less credible. On the other hand, journalists and media brands that are open about their use of AI tools may ultimately gain trust by demonstrating accountability and ethical practice.[3]

2. Huo Jingnan, It's quick and easy to clone famous politicians' voices, despite safeguards, May 31, 2024, National Public Radio https://www.npr.org/2024/05/30/nx-s1-4986088/deepfake-audio-elections-politics-ai

3. Konrad Collao, Ok Computer? Public Attitudes to the uses of Generative AI in News, July 16, 2024, Reuters Institute for the Study of Journalism https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2024-07/RISJ%20-%20OK%20Computer%20-%20News%20and%20AI%20-%20Report%20from%20CRAFT.pdf



Legal Risks

The use of GenAI in journalism introduces significant legal risks, many of which stem from the opaque nature of the technology, its terms of service, and the ambiguous status of its outputs. In relying on GenAI models, journalists and news outlets need to recognize the reputational and legal risks involved in unintentionally publishing biased, offensive, or copyrighted content.[4] An important point is that journalists and news organizations using tools like ChatGPT are legally responsible for the accuracy of AI-generated content. Providers such as OpenAI explicitly disclaim liability in their terms of use, instead putting accountability fully on the user. Therefore, if false or defamatory information produced by an AI system is inadvertently published, it could lead to libel or defamation claims against the journalist or outlet. Strategic lawsuits against public participation (SLAPPs) pose a particular threat to journalism. Additionally, AI-generated outputs may plagiarize or closely replicate proprietary content from other sources, raising the risk of copyright infringement. This is particularly problematic given that media organizations' own content may be scraped to train AI models without consent, compensation, or attribution, yet the copyright status of AI-generated content remains legally unsettled. Privacy breaches also present legal challenges: if sensitive or personally identifiable information is input into AI systems and later leaked or retrieved by malicious actors, it may violate data protection regulations. Compounding these risks is the lack of understanding transparency around how AI systems are trained and operate, making it difficult for journalists to establish the provenance of AI-generated material.

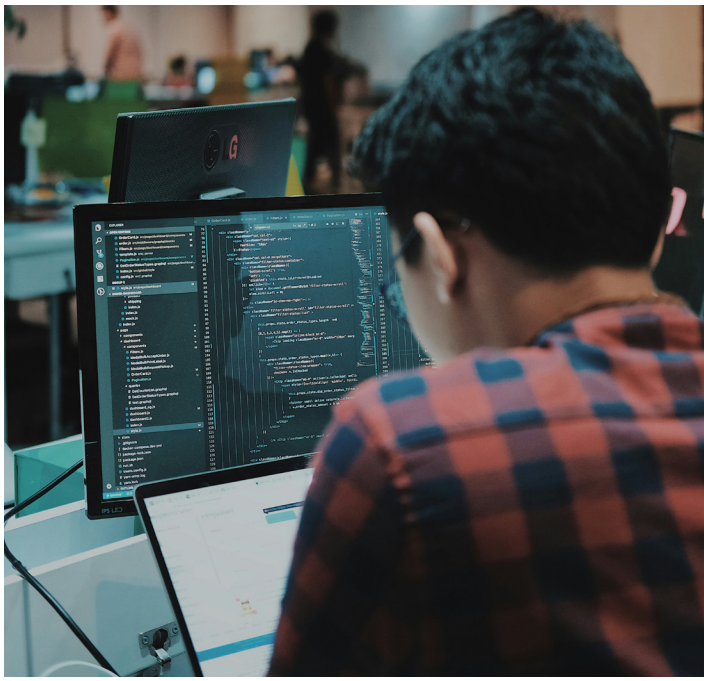
Amplifying these risks is the rapid pace of change. The legal landscape around GenAI is uncertain and fast evolving, with significant variations across jurisdictions, particularly in areas such as data protection, copyright, and defamation law. Smaller news organizations and individual journalists are especially vulnerable, lacking the bargaining power or legal resources to challenge the terms of service or copyright infringement, or to mitigate cross-border jurisdictional risks effectively. They must also be mindful of the fact that what may be legal in one country could violate the law in another. This is especially relevant when publishing for an international audience.

4. McKinsey & Company, *What is Generative AI*, 2024 https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai

CASE STUDY: FREEDOM OF INFORMATION (FOI) LAWS AND GENAI

In March 2025, in a significant development regarding transparency and the use of AI in government, the UK Department for Science, Innovation and Technology (DSIT) responded to a journalist's Freedom of Information (FOI) request concerning the ChatGPT interactions of the Secretary of State for Science and Technology, Peter Kyle. The request sought records of his communications with ChatGPT, including prompts and responses, conducted in his official capacity. DSIT complied and provided the requested records. This case sets a precedent for the application of FOI laws to AI-generated content, raising important questions about transparency and accountability in the digital age.

Michael Savage, Release of technology secretary's use of ChatGPT will have Whitehall sweating, March 13, 2025, The Guardian
https://www.theguardian.com/politics/2025/mar/13/peter-kyle-technology-secretary-freedom-of-information-chatgpt



Digital Risks

GenAI introduces complex and evolving digital risks for journalists and news outlets, especially as it becomes more embedded in day-to-day workflows. In that context, the most obvious risk for journalists relates to data privacy and leakage, which can happen when journalists inadvertently input proprietary, sensitive, or personally identifiable information into GenAI tools, leading to unintended data exposure. For example, the inclusion of names, locations, or other identifiers runs the risk of compromising individuals or sources, which can be particularly dangerous in closed, fragile, or high-surveillance environments. Hostile actors, including governments, corporations, and nonstate actors, may also exploit GenAI tools to conduct surveillance or extract compromising data. Also, GenAI could potentially be used to design more effective spear-phishing campaigns against journalists or outlets through automating and enhancing key aspects of social engineering.

Breached systems can result in an exposure of contacts, story leads, and whistleblower identities, threatening journalistic integrity and physical safety. Because many GenAI tools don't require users to sign in, they're easy to access—but this convenience can come at a cost. Users of free-tier tools may unknowingly give up control over their data, privacy, or even the rights to the content they create. This creates potential liabilities if content is reused, retained, or shared in ways that violate confidentiality agreements or data protection laws. The reality is that the rapid development of GenAI has outpaced most regulatory frameworks. And, in parallel, the competitive rush to dominate the AI market has led to the rapid release of tools and features—often without thorough checks, safeguards, or regulatory oversight. For journalists and media organizations, this creates a high-stakes environment where critical infrastructure is dependent on opaque, externally governed

systems, heightening vulnerabilities related to data privacy, manipulation, and misuse. And without clear protections in place, missteps in data handling, especially involving confidential sources, can have far-reaching consequences.

The security risks associated with GenAI are particularly acute for women journalists, who are already disproportionately the targets of online violence.[5] GenAI has exacerbated forms of technology-facilitated violence, including the use of AI-generated deepfakes to create nonconsensual sexual images intended to humiliate, silence, or discredit women in the public sphere.Such tactics not only harm their professional standing and personal well-being but also pose serious threats to media freedom, as they aim to push women out of the media space through intimidation and reputational damage. As regulation evolves to keep pace with new digital threats, some countries have made it a criminal offense to manipulate or distribute nonconsensual sexual images. Given that laws vary by jurisdiction, journalists have greater protection, both legally and personally, when operating in countries where such abuse is explicitly criminalized.

5. ICFJ, UNESCO, Online violence against women journalists: a global snapshot of incidence and impact, 2023 https://unesdoc.unesco.org/ark:/48223/pf0000375136

Practical Guidance for Media Outlets and Journalists

This section sets out practical, principle-based guidelines to help journalists and media outlets use GenAI responsibly, in line with IREX's and Development Gateway's commitment to RAI.

While the guidance provided here is structured into two sections—tailored to institutional and individual level risks, respectively—a single overarching principle applies to both: the essential role of human oversight at every stage of the journalistic process. Critically, GenAI should never be relied upon to make editorial decisions or to generate reporting without meaningful human oversight. These tools are best used as a complement to human editorial judgment and should remain aligned with established ethical guidelines and journalistic values.

Institutional Level—Guidance for Media Outlets

For media outlets, adapting to GenAI isn't just a technological shift but also an organizational and cultural one. Integrating it requires thoughtful implementation to ensure journalistic integrity and retain audience trust.

What follows is practical guidance for media outlets to ensure the safe, transparent, and responsible use of GenAI, and one that will need to be updated on an ongoing basis in line with technical and regulatory developments.

DEVELOP A GENAI POLICY

It is important for media outlets to recognize that people working in their newsrooms are already likely to be using GenAI tools informally. Thus, each media outlet needs to proactively address this risk by establishing a clear, transparent, and organization-wide policy on the responsible use of GenAI, establishing clear standards for how the media outlet uses, adopts, and engages with GenAI tools.[6]

A responsible policy on GenAI integration should:

- *Align with ethical and editorial standards:* Ensure that the institutional level policy is grounded in the outlet's core values—such as accuracy, accountability, and a commitment to "do no harm."

- *Apply a risk-based approach:* Conduct a structured risk assessment to identify potential harms, define how they will be monitored, and determine mitigation strategies aligned with the newsroom's values.

- *Initiate a participatory consultation process:* Engage staff from across the newsroom in shaping the policy. A transparent process enhances relevance, fosters buy-in, and surfaces practical concerns from those closest to the work.

- *Prioritize learning and integration:* Prioritize employee learning and encourage cross-departmental collaboration with the aim of embedding GenAI thoughtfully across editorial and operational workflows.

- *Leverage existing resources:* Draw on established industry-specific resources and models to inform policy design, such as WIRED's *Guidelines for Generative AI Tools*[7] and Poynter Institute's *AI Ethics Policy Template*.[8]

ESTABLISH A GENAI TASK FORCE

An effective GenAI task force should comprise staff from different departments across the outlet—including editorial, legal, technical, and operational team members—to ensure broad cross-departmental integration. Allocated responsibilities could include:

- Monitoring and oversight of GenAI usage across the media outlet, ensuring that all applications align with ethical standards, editorial values, and the outlet's mission.

- Keeping abreast of developments in the rapidly evolving GenAI landscape and regularly reviewing and updating the organization's GenAI policy. For example, the

6. Development Gateway: An IREX Venture, *How to balance the opportunities and risks of generative AI*, 2024, apolitical https://apolitical.co/solution-articles/en/how-to-balance-the-opportunities-and-risks-of-generative-ai

7. *How WIRED Will Use Generative AI Tools*, May 22, 2023, WIRED https://www.wired.com/about/generative-ai-policy/

8. Kelly McBride, *Your newsroom needs an AI ethics policy. Start here.* March 25, 2024, Poynter https://www.poynter.org/ethics-trust/2024/how-to-create-newsroom-artificial-intelligence-ethics-policy/?utm_source=chatgpt.com

Associated Press revisits its guidelines every three months.[9]

- Driving education and coordination across teams, and fostering thoughtful and integrated adoption of GenAI tools across the organization.
- Advising on broader actions related to GenAI, for example, whether to invest in customized GenAI systems or enter into licensing agreements with AI developers.

PUBLICLY DEFINE A GENAI POLICY

By being transparent about how they use GenAI, media outlets can reinforce public trust in ethical standards and journalistic integrity and also strengthen credibility with advertisers.

A public statement on GenAI should include:

- Transparency: Communicate how the media outlet will—and won't—use GenAI.[10]
- Accountability: Clarify that the media outlet is fully responsible and accountable for the content it produces, regardless of the role AI might play in content creation.
- Human oversight: Highlight a commitment to human oversight at every stage of the journalistic process.
- Ethical innovation: Demonstrate a commitment to maintaining journalistic authenticity while embracing new technologies.
- Continuous evaluation: Highlight how the outlet will regularly update policy to keep pace with technological advancements.

INVEST IN GENAI LITERACY

Investing in GenAI literacy, both internally and externally, is critical for upholding journalistic integrity and maintaining public trust.

Internally:

- Roll out outlet-wide training, providing staff, contractors and regular freelancers employees with a solid grounding in how GenAI works, including both its capabilities and limitations.
- Build outlet-wide awareness of GenAI policy and related protocols (for example, how to label AI-generated text, images, and video; disclosures of AI-assisted analysis; etc.).
- Invest in the critical evaluation skills of editors and journalists, equipping them with the skills and tools they need to confidently assess, fact-check, and verify AI-generated outputs, and foster editorial judgment over AI-generated suggestions. Ongoing training will further reinforce responsible use.

9. David Bauder, AP, other news organizations develop standards for use of artificial intelligence in newsrooms, August 17, 2023, Associated Press https://apnews.com/article/artificial-intelligence-guidelines-ap-news-532b417395df6a9e2aed57fd63ad416a

10. Examples of public statements by outlets such as the *New York Times, Financial Times*, Associated Press, and others can serve as useful templates when developing this statement.

Externally:

- Introduce an externally facing "Our Newsroom and AI" hub that outlines the media outlet's policy on GenAI, along with more general FAQs, explainers, and resources. This hub should be regularly updated to reflect most current developments.
- Demonstrate an ongoing commitment to clearly communicating to the public how AI is used in content creation, including labeling and issuing disclosures across all journalism. This transparency builds trust.
- Invest in technology reporting that explains and contextualizes GenAI, demystifying it for audiences, and thereby fostering greater public understanding and audience trust.

STRENGTHEN EDITORIAL PROCESSES

It is imperative that editorial processes in the age of GenAI remain firmly rooted in journalism's core principles: accuracy, independence, and integrity, with safeguards for in place for human oversight at every stage.

- Encourage a culture of critical engagement with GenAI. GenAI tools should be approached as enablers of journalism—supporting rather than supplanting journalistic judgment.
- Develop clear editorial policies and workflows for AI use, notably including robust content review processes for all AI-assisted outputs that ensure human oversight at every stage, from pitch to publication.
- Establish dual internal lines of responsibility. The onus is on editors to apply rigorous journalistic standards and meticulously fact-check, and journalists have a responsibility to be transparent about when, where, and how they used GenAI in their reporting processes. Embed safeguards from pitch to publication, for example, maintaining records of prompts and prompt chains to ensure transparency and traceability.
- Consider using tools to adjust to a desired tone or prompts that build unique voice profiles or voiceprints to mitigate the risk of homogenization of content.
- Clearly disclose when AI tools have been used in research or reporting, in line with the outlet's GenAI policy.

PRIORITIZE SAFETY

GenAI introduces complex and evolving digital risks. A strong understanding of these risks is essential to protect newsroom operations, uphold journalistic standards, and ensure the safety of both staff and sources:

- Establish and enforce clear digital safety protocols, including escalation procedures for suspected breaches or misuse, with safeguards in place to prevent harm.
- Evaluate the risk profile of different use cases, and choose GenAI tools accordingly.
- Apply a Data Sensitivity Rule that information too sensitive to publish should only be used with GenAI tools that have appropriate security and privacy protocols.

- Implement data masking and pseudonymization techniques, de-identify or anonymize all inputs for GenAI prompts to protect identities and minimize risk, and remove confidential business information.

- Address the growing threat of deepfake abuse, particularly against women journalists who are disproportionately the targets of online violence to undermine their reputation and their reporting..[11]

- Consider investing in customized AI systems where appropriate, which limit data exposure and operate on proprietary, secure data sets. Donors may be open to consider covering the costs of GenAI enterprise licenses as part of broader grant funding to nonprofit media outlets.

- Designate a focal point within the AI Task Force to lead on GenAI safety policy, monitor updates, and adapt organizational safeguards as the technology evolves.

11. Julie Posetti, *The Chilling: global trends in online violence against women journalists*, 2021 UNESCO https://unesdoc.unesco.org/ark:/48223/pf0000377223

Individual Level—Tailored Guidance for Journalists

There is no one-size-fits-all solution for the use of GenAI use in journalism. While the core principles of safety apply to both media outlets and individual journalists, the context and capacity differ sharply.

For freelance journalists and individual digital creators juggling multiple roles, GenAI offers invaluable support. These journalists often serve as their own editors, researchers, and marketers. GenAI tools can help them by summarizing documents, generating interview summaries, helping with grant writing, or drafting content such as pitch letters. Without access to copyeditors or fact-checkers, freelancers can use GenAI to perform first-pass edits, suggest alternative phrasings, or simulate interview questions to sharpen their reporting. The affordability of GenAI makes it especially valuable, offering low- or no-cost support for tasks that might otherwise require costly subscriptions or external help. Indeed, to quote an oft-cited mantra, think of GenAI as "the smartest personal assistant in the world."

However, all these positives are accompanied by risk factors. Individual journalists navigate these risks independently, often relying on public GenAI tools, personal judgment, and self-education, but without access to the latest threat intelligence or training tools. These journalists are thus more vulnerable to risks like data exposure, misinformation, and ethical missteps.

The absence of institutional oversight means that individual journalists must be vigilant in their use of GenAI; ensure alignment with ethical, professional, and legal standards; don't fall prey to misinformation; and don't inadvertently expose sensitive data. The reality is that they must take personal responsibility for safe and ethical GenAI use—because, quite simply, the risks of not doing so are too great.

DATA CONFIDENTIALITY

- Exercise caution inputting sensitive information in free-tier GenAI prompts. Aim to avoid including source identities, details of unpublished investigations, or anything that could put someone at risk.

- Implement data masking and pseudonymization techniques, and de-identify or anonymize content when referencing real-world individuals or situations. For example, use placeholder names or general references.

- Have a clear understanding of how data are retained or used according to the AI tool's policy.

DIGITAL SECURITY

Always review the terms of service and data policies of any GenAI tool you use. When possible, use incognito or private modes to keep your prompts out of history, and choose tools that don't use your data to train models by default.

Be cautious of using a browser-based tool. Ideally, use only trusted, privacy-conscious platforms when working on sensitive material. For example, opt for VPNs or trusted enterprise-level accounts, whenever possible.

- Enable multifactor authentication and follow strong password practices on GenAI accounts.

- Avoid working with sensitive materials on personal or unsecured devices.

ONLINE VIOLENCE

- Familiarize yourself with the emerging risks around online violence and be aware of how GenAI can be used maliciously for harassment, disinformation, or deepfake creation.

- If you experience or witness GenAI-enabled harassment, report it—either to the outlet publishing your work, to the police in countries where manipulating or distributing nonconsensual sexual images is a criminal offense, to relevant safety networks, or to professional associations. A helpful resource for women journalists is the Online Violence Response Hub.[12]

- Be aware of the mental health toll of online violence. Consider using a mental health self-evaluation chart so you can assess how online violence is affecting your well-being.[13]

ETHICAL USE AND EDITORIAL OVERSIGHT

- Be wary of overreliance. GenAI should complement, not replace, journalistic judgment.

- Any output from a GenAI tool should be treated as unvetted source material. Independently verify and cross check all AI-generated outputs.

- Document how you use AI tools in your reporting workflow. For example, some tools like Claude and ChatGPT let you get a public link for individual prompt threads to help you track and cite your work with gen AI, and which can be provided to editors to facilitate fact-checking and traceability.

- When collaborating with editors or publishers, communicate your use of AI transparently and follow any applicable guidelines.

LEGAL

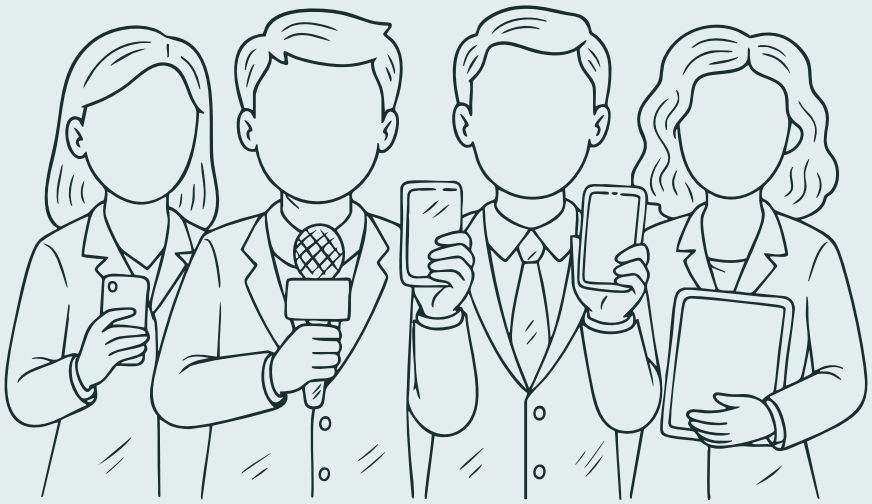
- Stay informed about defamation, copyright, and privacy laws in your region and those relevant to your reporting, to ensure your use of GenAI aligns with applicable laws and professional norms.

- Draw on the support of networks such as the Legal Network for Journalists at Risk (LNJAR),[14] Journalists in Distress (JID) Network[15] as well as journalist support organizations focused on legal matters such as Media Defence[16] to help you navigate legal issues.

12. Coalition Against Online Violence, Online Violence Response Hub https://onlineviolenceresponsehub.org/
13. Ana Maria Zellhuber Pérez & Juan Carlos Segarra Pérez, A Mental Health Guide for Journalists facing Online Violence, 2022, IWMF https://www.iwmf.org/wp-content/uploads/2022/12/Final_IWMF-Mental-health-guide.pdf
14. https://www.medialegalhelp.org/about-the-network/
15. https://www.journalistsindistress.org
16. https://www.mediadefence.org/

ONGOING TRAINING AND AWARENESS

- Ultimately, individual journalists must build what's known as "AI agency"— the ability to critically assess, safely use, and responsibly integrate GenAI into their workflows.
- Stay informed about evolving capabilities and risks, notably in terms of issues like hallucinations, deepfakes, and dis- or misinformation.
- Seek out tailored resources and training opportunities, to learn how to work with GenAI tools safely, ethically, and effectively.



GenAI is still a very new field, and the landscape of risks and opportunities continues to evolve rapidly. As GenAI becomes increasingly incorporated across the media sector, and with a greater shift toward agentic AI,[17] careful management is required of the evolving safety and security risks to protect journalists, sources, audiences, and institutional integrity. As journalists and media outlets begin experimenting with these tools, it is crucial to keep a finger on the pulse of regulation and risk.

Realizing the significant benefits of GenAI for media and journalism will require careful integration guided by public interest values—ensuring it supports informed societies, strengthens trust in information, and upholds democratic discourse. With the right safeguards, GenAI can become a powerful tool to enhance newsroom efficiency, expand access to information, and foster more inclusive and responsive journalism.

The more long-term question perhaps is what being exposed to computer-generated content does to people's overall sense of trust in news reports. Both disclosing the use of AI and using AI without transparency have the potential to undermine public trust in organizations and potentially have wider ripple effects to erode trust in the media sector.[18] This is a significant societal risk and one that needs to be kept to the forefront of the discussion around the use of GenAI in media.

As Nic Newman of The Reuters Institute for Journalism has noted, public perception will play a crucial role in shaping the adoption of AI in the news industry. While awareness of AI tools is growing rapidly, trust and acceptance remain significant hurdles. The speed at which AI is adopted and how institutions implement it will be critically defined by how audiences respond.[19]

17. Marina Adami, *Nordic AI in Media Summit 2025: Five takeaways from this annual event on the future of news*, April 25, 2025, Reuters Institute for the Study of Journalism https://reutersinstitute.politics.ox.ac.uk/news/nordic-ai-media-summit-2025-five-takeaways-annual-event-future-news

18. Benjamin Toff & Felix M. Simon, "Or They Could Just Not Use It?" The Dilemma of AI Disclosure for Audience Trust in News, December 2024, Sage Journals https://journals.sagepub.com/doi/abs/10.1177/19401612241308697

19. Carlo Prato The Digital Growth Summit 2024, October 9, 2024, Twipe https://www.twipemobile.com/the-digital-growth-summit-2024-in-10-quotes/

Acknowledgments

This guide has drawn on the experiences, insights, and lessons learned from our work with media partners around the world and others who have published on the topic. Many IREX and Development Gateway staff members provided invaluable knowledge, expertise, and insights that enabled this guide to come to fruition. In particular, this resource draws on previous work by Tara Susman-Pena and Pavle Zlatić and has benefited from significant contributions from Samhir Vasdev, Jill Miller, and Josh Powell. GenAI was used for the initial copyediting of this document. All outputs were reviewed, evaluated, and edited by the lead authors.

Authors
Mary O'Shea, Senior Technical Advisor – Information & Media, IREX
Tom Orrell, Deputy Director of Programs, Development Gateway

Designer
Sebastián Molina

About IREX
IREX is a global development and education organization. We strive for a more just, prosperous, and inclusive world in which individuals reach their full potential, governments serve their people, and communities thrive. We work with partners in more than 100 countries in four areas essential to progress: cultivating leaders, empowering youth, strengthening institutions, and increasing access to quality education and information. Learn more: www.irex.org.

About Development Gateway: An IREX Venture
Development Gateway is an international nonprofit that uses digital technology and evidence to create more effective, responsive, and trusted institutions. It supports digital and data solutions for development, bringing technology and a programmatic lens to each project. In doing so, the Development Gateway helps governments and civil society to strengthen data and digital governance and take a strategic approach to using digital for good. Learn more: www.developmentgateway.org

©2025. This resource is openly licensed under CC BY-SA 4.0. This means you are welcome to use and adapt it so long as you give appropriate credit, indicate if you make changes, and distribute under the same license.

Annex 1: Glossary

acquired data
Information purchased or licensed from third parties, often collected through partnerships, agreements, or commercial data brokers.

anonymization
Process of removing personally identifiable information from data sets or content to protect the privacy of individuals involved.

artificial intelligence (AI)
A field of computer science that creates systems capable of performing tasks that typically require human intelligence, such as pattern recognition, decision-making, or language generation.

bias in AI
Systematic errors in AI outputs that arise from biased training data, model design, or deployment processes.

black box system
An AI model whose internal decision-making processes are not visible, understandable, or easily explained to users, regardless of whether the model is closed or open, making it difficult to trace how an output was generated.

closed model system
An AI model whose training data, algorithms, and processes are proprietary and not disclosed to the public.

customized models
AI models that are tailored or fine-tuned to meet the specific needs or requirements of an organization, industry, or task; trained on or adjusted using specialized data.

data privacy
Ethical and legal obligation to protect individuals' personal or sensitive information when using digital tools, including AI systems.

deepfake
Synthetic media—usually video, images, or audio—in which a person's likeness or voice is manipulated using AI technologies to create deceptive or false representations.

de-identify
Process of removing or altering personally identifiable information (PII) from data to ensure privacy and confidentiality and reducing data protection risks.

editorial oversight
Processes through which newsroom leaders and editors review journalism to ensure it meets ethical, factual, and professional standards.

enterprise licenses
GenAI tools under customized contracts, which include stronger data protections, security guarantees, and compliance with privacy standards. These licenses are negotiated on a case-by-case basis, with pricing depending on the number of users, scale of the deployment, features required, and level of support needed.

ethical AI use
Practices that ensure AI technologies are deployed responsibly, respecting journalistic integrity, human rights, and societal values.

generative AI (GenAI)
A type of AI that creates new content—such as text, images, audio, or video—by learning from large data sets.

hallucination
When a GenAI produces information that is factually incorrect, fabricated, or nonsensical while appearing plausible and confident.

prompt chain
Sequence of related prompts and GenAI responses. Tracking prompt chains helps ensure transparency, reproducibility, and editorial oversight in AI-assisted work.

prompt injection
A type of attack or manipulation in GenAI systems where a bad actor intentionally alters the training data in a way that causes the model to produce unintended or malicious outputs.

responsible AI (RAI)
Development and use of AI in a manner that prioritizes safety, fairness, accountability, transparency, and respect for human rights. It emphasizes minimizing risks and harms while maximizing the positive impact of AI technologies.

scraped data
Information automatically collected from public websites or online platforms, typically without direct permission from content owners.

source transparency
The practice of clearly disclosing where information comes from, including the use of AI tools in content creation.

tech-facilitated violence
Harmful actions such as harassment, threats, or abuse carried out through digital technologies.

terms of service (TOS)
Legal agreement between users and service providers that outlines how a digital platform or AI tool can be used, including data usage and ownership rights.

training data
Data sets used to teach an AI model to recognize patterns, predict outcomes, or generate content, and can include public, private, and proprietary materials.

Annex 2: Key Applications: How Generative AI Is Changing the News Media

The impact of GenAI on the media sector has been profound. From automated content creation to personalized news delivery, GenAI is reshaping how stories are produced, distributed, and consumed. Indeed, in a short space of time, it has gone from being a novelty to a necessity in newsrooms, transforming how media and journalism operates, with powerful applications across operational and editorial spheres.

When used thoughtfully, GenAI can enhance both creativity and workflow, unlocking efficiencies, reaching new audiences, enabling personalization, and advancing the mission of journalism in the digital age.

KEY APPLICATIONS:

Research Support:

GenAI helps journalists work faster, sift through more information, and spot key insights more easily, greatly accelerating the journalistic research process. It does so by summarizing large volumes of information, extracting key quotes, translating non-English sources, or organizing information. For example, deep research models from Gemini, ChatGPT, and others can perform detailed and well-sourced analyses about topics, freeing up time for journalists to spend on deeper analysis.

Content Creation:

In a fast-moving news cycle, GenAI helps draft articles, headlines, and summaries quickly, freeing up time for deeper reporting and analysis. It enables outlets to develop unique voiceprints that allow gen AI tools to produce content using tone, language, and style that matches their brand and audience, as well as adapting content for different platforms, helping to transform information into compelling narratives—whether visual, interactive, or immersive—improving reach and engagement.

Dynamic Visuals:

GenAI empowers journalists to produce visually compelling content, critical in an increasingly digital and visual media landscape. Infographics, charts, videos, maps, and animations—formats more likely to capture attention and drive engagement—help simplify complex information, making reporting easier to understand and more impactful.

Audience Engagement:

GenAI expands the modes of expression available to journalists. In creating formats more likely to capture attention and drive engagement, reporting can reach wider and more diverse audiences. In conjunction, AI-powered personalization (predictive AI) is also powering engagement across social

media channels, opening new avenues for strengthening audience connection.

Workflow Efficiency:

By automating routine and time-consuming tasks, GenAI is bolstering a faster turnaround and reduced workload and thus freeing up more time for quality reporting and analysis. Such efficiency is especially critical in fast-paced news environments. Examples of AI tools for transcription and summarization are Firefly or Otter.ai, which can be used during interviews or when digitizing information like photographs or spreadsheets.

Innovation and Experimentation:

GenAI enables rapid prototyping of new content formats and storytelling methods, making it easier for small or resource-constrained teams to experiment and scale production without large investment.

Accessibility:

When used thoughtfully, GenAI can significantly broaden who journalism reaches and how effectively it communicates. Inbuilt features make content more understandable, inclusive, and usable for a wider range of audiences. Examples include creating voiceovers, captions, or summaries for video/audio content for people with visual or hearing impairments or translations into other languages.

Business Development:

For nonprofit newsrooms, securing grants is essential to long-term sustainability. GenAI helps streamline the resource-intensive activity of business development, empowering those with limited time and limited staff to compete for philanthropic funding in a competitive field.

IREX

DEVELOPMENT GATEWAY
An IREX Venture

Guidance Note for Newsrooms & Journalists

Safety Considerations for the Use of Generative Artificial Intelligence (GenAI)

June 2025