Expert Perceptions of Generative-AI–Driven Disinformation

Purpose: This survey explores expert perceptions of generative-AI–driven disinformation for an academic research project. All responses will be treated as confidential and reported in an anonymised, aggregated format unless explicit consent is given for attribution.

Consent & Data Protection

Detailed Information & Data Protection Notice (GDPR)

By participating in this survey, you agree to the following terms regarding the collection and use of your data.

Data Controller: The person responsible for your data is Alexander Loth, conducting research at the Frankfurt University of Applied Sciences. Contact: `alexander.loth@stud.fra-uas.de`.

Purpose and Lawful Basis: The data from this survey will be used exclusively for academic research purposes, including analysis, publication in academic journals, and conference presentations. The legal basis for processing your personal data is your explicit consent, which you provide by starting the survey.

Data Handling: Your survey responses will be stored securely. By default, all published results will be fully anonymised. If you explicitly consent at the end of the survey, your name and affiliation may be included in the acknowledgements or used for direct attribution of quotes.

Data Retention: Anonymised research data will be archived indefinitely as part of the scientific record. Any personal data linking you to your responses (such as your name or email if provided for attribution) will be securely deleted one year after the final publication of this research, or immediately upon your request.

Your Rights as a Participant: In accordance with GDPR, you have the right to:

- **Withdraw Consent:** You can withdraw your consent at any time without penalty by closing the survey or by contacting the researcher via email to have your data deleted.
- **Access, Rectify, and Erase Data:** You have the right to request access to, correction of, or deletion of your personal data.
- **Lodge a Complaint:** You have the right to lodge a complaint with a supervisory authority. The responsible authority for the Frankfurt University of Applied Sciences is the Hessian Commissioner for Data Protection and Freedom of Information (*Hessischer Beauftragter für Datenschutz und Informationsfreiheit*).

By proceeding, you confirm that you have read and understood these terms.

* *I confirm I am 18 years or older and I have read, understood, and agree to the consent terms outlined above.*

◯ Yes, I consent to participate

◯ No, I do not consent

Section 1/9: Screening & Expert Group

* *Which best describes your primary professional role?*

◯ AI researcher / ML engineer (development of generative models)

◯ Disinformation / fact-checking professional

◯ Journalist covering tech or politics

◯ Policymaker / regulator involved with AI, media or digital-services policy

◯ Ethicist / legal scholar working on AI or deepfakes

◯ Other: _____

* *Please provide a link to a recent activity of yours that might be interesting for our research (e.g., recent paper title, organisational role, or link).*

Section 2/9: Demographics & Baseline Knowledge

* *Years of professional experience in your field.*

* *Region of primary professional activity.*

◯ European Union (EU)

◯ Europe (non-EU, including UK)

◯ North America (USA & Canada)

◯ East & Southeast Asia

◯ South & Central Asia

◯ Latin America and the Caribbean

◯ Middle East & North Africa (MENA)

◯ Sub-Saharan Africa

◯ Oceania (Australia, New Zealand, etc.)

◯ Other: _____

* *Self-rated technical understanding of Large Language Models.*

Likert scale: 1 (Novice) – 7 (Expert)

1	2	3	4	5	6	7
◯	◯	◯	◯	◯	◯	◯

* *Self-rated familiarity with deepfake video technology.*

Likert scale: 1 (Novice) – 7 (Expert)

1	2	3	4	5	6	7
◯	◯	◯	◯	◯	◯	◯

Section 3/9: Perceived Threats (Modality Only)

Q9. *On a 1–7 scale (1 = Not a threat, 7 = Severe threat), how threatening is each modality for spreading disinformation in general?*

Modality	1	2	3	4	5	6	7
AI-generated Text	◯	◯	◯	◯	◯	◯	◯
AI-generated Images	◯	◯	◯	◯	◯	◯	◯
AI-generated Audio (voice cloning)	◯	◯	◯	◯	◯	◯	◯
AI-generated Video (deepfakes)	◯	◯	◯	◯	◯	◯	◯

Section 4/9: Perceived Threats (Modality × Domain Matrix)

Now consider specific domains. How threatening is each modality in the given domain?

Q10. *Threats in the **Political** Domain (1 = Not a threat, 7 = Severe threat).*

Modality	1	2	3	4	5	6	7
AI-generated Text	◯	◯	◯	◯	◯	◯	◯
AI-generated Images	◯	◯	◯	◯	◯	◯	◯
AI-generated Audio	◯	◯	◯	◯	◯	◯	◯
AI-generated Video	◯	◯	◯	◯	◯	◯	◯

Q11. *[Optional] Which specific points concerning the Political Domain would you like to share with us?*

Q12. *Threats in the **Health** Domain (1 = Not a threat, 7 = Severe threat).*

Modality	1	2	3	4	5	6	7
AI-generated Text	◯	◯	◯	◯	◯	◯	◯
AI-generated Images	◯	◯	◯	◯	◯	◯	◯
AI-generated Audio	◯	◯	◯	◯	◯	◯	◯
AI-generated Video	◯	◯	◯	◯	◯	◯	◯

Q13. *[Optional] Which specific points concerning the Health Domain would you like to share with us?*

Q14. *Threats in the **Financial** Domain (1 = Not a threat, 7 = Severe threat).*

Modality	1	2	3	4	5	6	7
AI-generated Text	◯	◯	◯	◯	◯	◯	◯
AI-generated Images	◯	◯	◯	◯	◯	◯	◯
AI-generated Audio	◯	◯	◯	◯	◯	◯	◯
AI-generated Video	◯	◯	◯	◯	◯	◯	◯

Q15. *[Optional] Which specific points concerning the Financial Domain would you like to share with us?*

Q16. *Threats in the **Social** Domain (1 = Not a threat, 7 = Severe threat).*

Modality	1	2	3	4	5	6	7
AI-generated Text	◯	◯	◯	◯	◯	◯	◯
AI-generated Images	◯	◯	◯	◯	◯	◯	◯
AI-generated Audio	◯	◯	◯	◯	◯	◯	◯
AI-generated Video	◯	◯	◯	◯	◯	◯	◯

Q17. *[Optional] Which specific points concerning the Social Domain would you like to share with us?*

Section 5/9: Key Risks (Top-3)

* *Select up to three risks you judge most urgent.*

- ☐ Election interference via deepfake video
- ☐ Large-scale text spambots (astroturfing)
- ☐ Audio-clone financial fraud / scams
- ☐ Synthetic medical misinformation
- ☐ Harassment / non-consensual deepfake porn
- ☐ Cyberattacks (e.g. phishing)
- ☐ Other: _____

Q18. *[Optional] In 1–2 sentences, tell us why the risks you selected worry you most. Feel free to cite a real incident or scenario.*

Section 6/9: Mitigation Strategies – Ratings

Q20. *On a scale of 1 (Not effective) to 7 (Very effective), rate the effectiveness of the following mitigation strategies.*

Strategy	1	2	3	4	5	6	7
Technical detection tools	◯	◯	◯	◯	◯	◯	◯
Digital watermarking / provenance standards	◯	◯	◯	◯	◯	◯	◯
Government regulation (e.g., AI Act, DSA)	◯	◯	◯	◯	◯	◯	◯
Media / public literacy programmes	◯	◯	◯	◯	◯	◯	◯
Platform enforcement policies	◯	◯	◯	◯	◯	◯	◯

Q21. *[Optional] Which additional information concerning mitigation strategies would you like to share with us?*

Section 7/9: Mitigation Strategies – Forced Ranking

Q22. *Please rank the following strategies from 1 (most effective) to 5 (least effective).*

Strategy	1	2	3	4	5
Technical detection tools	◯	◯	◯	◯	◯
Digital watermarking / provenance standards	◯	◯	◯	◯	◯
Government regulation (e.g., AI Act, DSA)	◯	◯	◯	◯	◯
Media / public literacy programmes	◯	◯	◯	◯	◯
Platform enforcement policies	◯	◯	◯	◯	◯

Section 8/9: Best–Worst Scaling (Preference Weights)

* *Which strategy is **most** effective?*

- ◯ Digital watermarking / provenance standards
- ◯ Media / public literacy programmes
- ◯ Government regulation (e.g., AI Act, DSA)
- ◯ Technical detection tools
- ◯ Platform enforcement policies

* *Which strategy is **least** effective?*

- ◯ Digital watermarking / provenance standards
- ◯ Media / public literacy programmes
- ◯ Government regulation (e.g., AI Act, DSA)
- ◯ Technical detection tools
- ◯ Platform enforcement policies

Q23. *[Optional] What is the single biggest obstacle (technical, political or economic) that could block successful adoption of your top-ranked mitigation strategy?*

Section 9/9: Evaluating Mitigation Strategies – Public Literacy Tools

The next set of questions concerns a specific category of mitigation strategy: public-facing awareness tools designed to improve media literacy. For the following questions, please consider a specific, real-world example called JudgeGPT (`https://judgegpt.streamlit.app/`).

Methodology of Public Literacy Tools. Tools for public disinformation literacy typically have the following characteristics: Users are shown a series of short text-based news fragments. For each fragment, the user must make two separate judgments using rating sliders: (1) *Origin Detection:* Is the text written by a human or generated by an AI? (2) *Veracity Judgment:* Is the information in the news fragment legitimate or fake? After completing a set of five fragments, the application shows the user their performance statistics, including their accuracy scores for both tasks and awards badges for high performance.

Q26. *On a 1–7 scale, how effective do you believe such public-awareness tools are for mitigating AI-generated fake news?*

Likert scale: 1 (Not Effective) – 7 (Very Effective)

1	2	3	4	5	6	7
◯	◯	◯	◯	◯	◯	◯

* *Select the single greatest limitation you see in such tools.*

- ◯ Easily bypassed by adversarial content
- ◯ Reaches only tech-savvy audiences
- ◯ Impact fades once novelty wears off
- ◯ Resource-intensive to maintain
- ◯ Other: _____

Q27. *[Optional] What do you see as the main strength of public-awareness tools like JudgeGPT?*

Q28. *[Optional] Suggest one concrete feature or programme that would make public-resilience tools more effective.*

Future Outlook & Final Comments

Q30. *[Optional] Looking ahead five years, name one emerging or currently underestimated risk at the AI–disinformation frontier.*

Q31. *[Optional] Any final thoughts we haven't covered, or feedback on the survey?*

Attribution & Final Submission

As mentioned in the consent form, we would like to offer you the option to be acknowledged for your contribution. Please select one of the options below. The default is full anonymity.

* *How would you like your contribution to be handled in the final publication?*

- ◯ Please keep my responses and participation completely anonymous.
- ◯ You may list my name and affiliation in the paper's "Acknowledgements" section.
- ◯ In addition to acknowledging me, you may attribute specific quotes or ideas from my open-ended responses directly to me by name.

* *Full Name* _____

* *Affiliation* _____

Q32. *[Optional] Link to your preferred professional/academic profile.* _____