
User Transparency Within AI¹

Jonathan King
Howard University

Robert Hardy
Howard University

Jeremiah Bailey
Howard University

Amir Abdulgadir
Howard University

With
Howard University & Apart Research

Abstract

Generative AI technologies present immense opportunities but also pose significant challenges, particularly in combating misinformation and ensuring ethical use. This policy paper introduces a dual-output transparency framework requiring organizations to disclose AI-generated content clearly. The proposed system provides users with a choice between fact-based and mixed outputs, both accompanied by clear markers for AI generation. This approach ensures informed user interactions, fosters intellectual integrity, and aligns AI innovation with societal trust. By combining policy mandates with technical implementations, the framework addresses the challenges of misinformation and accountability in generative AI.

Keywords: Multi-agent alignment, AI security, model evaluations, safety infrastructure

1. Introduction

Generative AI technologies present immense opportunities but also pose significant challenges, particularly in combating misinformation and ensuring ethical use. This policy paper introduces a dual-output transparency framework requiring organizations to disclose AI-generated content clearly. The proposed system provides users with a choice between fact-based and mixed outputs, both accompanied by clear markers for AI generation. This approach ensures informed user interactions, fosters intellectual integrity, and aligns AI innovation with societal trust. By combining policy mandates with technical implementations, the framework addresses the challenges of misinformation and accountability in generative AI.

2. Methods

Generative AI technologies present immense opportunities but also pose significant challenges, particularly in combating misinformation and ensuring ethical use. This policy paper introduces a dual-output transparency framework requiring organizations to disclose AI-generated content clearly. The proposed system provides users with a choice between fact-based and mixed outputs, both accompanied by clear markers for AI generation. This approach ensures informed user interactions, fosters intellectual integrity, and aligns AI innovation with societal trust. By combining policy mandates with technical implementations, the framework addresses the challenges of misinformation and accountability in generative AI.

3. Results

The dual-output framework is designed to provide users with clear choices while holding organizations accountable. Fact-based outputs will be exclusively sourced from verifiable databases and include citations for transparency. Mixed outputs, integrating generative AI content, will carry distinct markers and disclaimers to inform users of their origins. Organizations must issue public statements detailing their AI disclosure practices, ensuring clarity in how content is generated and verified.

Implementation will proceed in three phases. First, technical specifications will be developed over six months, focusing on system compatibility and user experience. This will be followed by a pilot program with industry leaders, lasting one year. Finally, the policy will be rolled out industry-wide within three years, accompanied by compliance monitoring and periodic evaluations.

Success metrics include increased user trust in AI platforms, reductions in misinformation cases, and high adoption rates of transparency protocols. Initial pilot studies with simulated outputs suggest significant potential for improved user satisfaction and trust in AI systems.

4. Discussion

The proposed policy is expected to yield substantial benefits across short, medium, and long-term horizons. In the first two years, public awareness of AI-generated

content will improve as transparency markers become commonplace. Over three to five years, industry-wide adoption of disclosure practices will establish new norms for content accountability. In the long term, the policy will contribute to societal trust in AI technologies, ensuring their responsible integration into information ecosystems.

Stakeholders will experience diverse impacts. AI developers may face initial costs to implement dual-output systems, but these will be offset by enhanced user trust and reduced legal risks. Governments will gain a practical framework for enforcing AI transparency, while users will benefit from greater clarity and reduced exposure to misinformation. Potential challenges, such as resistance from developers and technical constraints, can be mitigated through phased implementation and stakeholder engagement.

The policy is both politically and economically feasible. Politically, it aligns with global efforts to regulate AI ethically, such as initiatives led by the European Union and UNESCO. Economically, the required infrastructure is compatible with existing AI capabilities, minimizing cost barriers for developers. Technically, dual-output systems are achievable using current generative AI models, ensuring scalability and practicality.

5. Conclusion

Generative AI is reshaping the way information is produced and consumed, but without robust transparency measures, its potential benefits are overshadowed by risks of misinformation and diminished trust. The proposed dual-output framework offers a balanced approach to these challenges, combining policy and technical solutions to enhance user trust and accountability. By implementing this framework, stakeholders can ensure that AI serves as a tool for innovation and integrity, rather than a source of misinformation. Immediate action is crucial to establish transparency as a cornerstone of AI governance.

6. References

- Goh, G., et al. "Multimodal Neurons in Artificial Neural Networks." *Distill*, vol. 6, no. 3, 2021, pp. 10–20. <https://doi.org/10.23915/distill.00030>.
- Krishna, V. "AI and Contemporary Challenges: The Good, Bad, and the Scary." *Journal of Open Innovation*, vol. 10, no. 1, 2024.
- Vinuesa, R., Azizpour, H., van der Lee, R., & Silva, T. (2020). The role of artificial intelligence in achieving the sustainable development goals. *Nature Communications*, 11(1), 1-10.

<https://doi.org/10.1038/s41467-020-17967-7>

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). Dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1-24.
<https://doi.org/10.1145/3411764.3445698>

Marcus, G. (2020). *Rebooting AI: Building artificial intelligence we can trust*. Pantheon Books.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.