PARTNERSHIP ON AI

Responsible Practices for Synthetic Media Case Study

Template for Framework Supporter Organizations

1 Organizational Background

1. Provide some background on your organization.

2 Framing Direct Disclosure at your Organization

1. Please elaborate on how your organization recommends providing Direct Disclosure (as defined in our Glossary for Synthetic Media Transparency Methods) to users/audiences.
2. Does your organization understand the goal of Direct Disclosure as specified in the PAI Framework: "to mitigate speculation about content, support resilience to manipulation or forgery, be accurately applied, and communicate uncertainty without furthering speculation" or does it have a different understanding?
3. What, if anything, does your organization believe is missing from this NIST taxonomy? Should it be added to a taxonomy of Direct Disclosure? If so, why?

   From NIST's Reducing Risks Posed by Synthetic Content:

   > The most commonly used techniques to *directly disclose* to the audience how AI was used in the content creation process include:
   >
   > - content labels (e.g., visual tags within content, warning labels, pre-roll or interstitial labels in video and/or audio, and typographical signals in text highlighting generated AI text with different fonts),
   > - visible watermarks (e.g., icons covering content indicating AI usage where the bigger the icon, the harder its removal), and
   > - disclosure fields (e.g., disclaimers and warning statements to indicate the role of AI in developing the content, and acknowledgments to provide more context to the AI contribution and credits to reviewers).

4. What criteria should [Builders, Creators, and/or Distributors] of synthetic media use to determine whether content should be Directly Disclosed?
5. Per the Framework, PAI recommends disclosing "visual, auditory, or multimodal content that has been generated or modified (commonly via artificial intelligence). Such outputs are often highly realistic, would not be identifiable as synthetic to the average person, and may simulate artifacts, persons, or events."

3 Real World, Complex Direct Disclosure Example

1. Provide a real-world example in which either: a) Direct Disclosure should have been applied, or b) Direct Disclosure was applied to a piece, or category, of content for which it was challenging to evaluate whether it warranted a disclosure. This could be because the threshold for disclosing was uncertain, the impact of such content was debatable, understanding of how it was manipulated was unclear, etc. Be sure to explain *why* it is challenging.
2. How was this piece/kind of content identified?
3. Was there any potential for reputational (e.g., negative impact on the organization's brand, products, etc.), societal (e.g., negative impact on the economy, etc.), or any other kind of harm from such content?
4. What was the impact of implementing, or not implementing, this Direct Disclosure? How would your organization assess such impact (studying users, via the press, other civil society, community reactions, etc.)? Did the disclosure mechanism mitigate the harm described in the previous question (3.3)?

5. Is there anything your organization believes either the Builder, Creator, or Distributor of the content should have done differently to support Direct Disclosure?
6. In retrospect, what, if anything, does your organization believe should have been done differently by the stakeholders identified in the previous question?
7. Were there any other policy instruments that should have been relied upon in deciding whether to, and how, to disclose the content? What external policies may have been helpful to supplement internal policies?
8. What might industry practitioners or policymakers learn from this example? How might this case inform best practices for Direct Disclosure across those Building, Distributing, and/or Creating synthetic media?

4 How Organizations Understand Direct Disclosure

1. What research and/or analysis has contributed to your organization's understanding of Direct Disclosure (both internal and external)?
2. Does your organization believe there are any risks associated with either OVER or UNDER disclosing synthetic media to audiences? How does your organization recommend navigating these tensions?
3. What conditions or evidence would prompt your organization to re-calibrate your answer to the previous question (4.2)? E.g., in an election year with high stakes events, your organization may recommend over labeling.
4. In the March 2024 guidance from the PAI Synthetic Media Framework's first round of cases, PAI wrote of an emergent best practice: "Creative uses of synthetic media should be labeled, because they might unintentionally cause harm; however, labeling approaches for creative content should be different, and even more mindfully pursued, than those for purely information-rich content."

   Does your organization agree? If so, how do you think creative content should be labeled? What is your organization's understanding of "mindfully pursued"? If your organization does not agree, why not?
5. Overall, what role(s) does your organization believe Builders, Creators, and Distributors play in directly disclosing AI-generated or AI-edited media to users?
6. How important is it for those Building, Creating, and/or Distributing synthetic media to all align collectively, or within stakeholder categories, on a singular threshold for:

   1) the types of media that warrant Direct Disclosure, and/or
   2) more specifically, a shared visual language or mechanism for such disclosure?

   Elaborate on which values or principles should inform such alignment, if applicable.

5 Approaches to Direct Disclosure, in Policy and Practice

1. What does your organization believe are the most significant sociotechnical challenges to successfully achieving the purpose of Directly Disclosing content at scale? (Refer to question 2.3 for reference to PAI's description of Direct Disclosure)
2. What goals should organizations be trying to accomplish when implementing Direct Disclosure? Does your organization believe Directly Disclosing ALL AI-generated or modified, as several policies are recommending, is useful in helping accomplish those goals?
3. Please share your organization's insight into how Direct Disclosure can impact:

    1) Accuracy
    2) Trustworthiness
    3) Authenticity
    4) Harm mitigation
    5) Informed decision-making
    6) Anything else we're missing that is relevant here

    NOTE: You can also discuss your understanding of the relationship between these concepts (for example, authenticity could impact trustworthiness, harm mitigation, etc.)

4. Does your organization believe there will be a tipping point to the liar's dividend (that people doubt the authenticity of real content because of the plausibility that it's AI-generated or AI-modified), why or why not? If yes, have we already reached out? How might we know if we have reached it?
5. As AI-generated media becomes more ubiquitous, what are some of the other important questions audiences should be asking in addition to "is this content AI-generated or AI-modified," especially as more and more content today has some AI-modification.
6. How can research help inform development of Direct Disclosure that supports user/audience needs? Please list out key open areas of research related to Direct Disclosure that, the answers to which, would support your organization's understanding of Direct Disclosure.

6 Media Literacy and Education

1. In the March 2024 guidance from the Synthetic Media Framework's first round of cases, PAI wrote of an emergent best practice: "Broader public education on synthetic media is required for any of the artifact-level interventions, like labels, to be effective"

    Does your organization agree? If so, why? Has your organization been working on "broader public education on synthetic media"? How (please provide examples)? If your organization does not agree, why not? What responsibility do Builders, Creators, or Developers (as defined in the Framework) have in educating users? What about civil society organizations?

2. What would you like to see from other institutions as it relates to improving public understanding of synthetic media? Which stakeholder groups have the largest role to play in educating the public (e.g., civic institutions; technology platforms; schools)? Why?
3. What support does your organization need in order to advance synthetic media literacy and public education on evaluating trustworthiness?

OPTIONAL

7 Commentary on the Framework's first set of cases (beyond Direct Disclosure)

1. The first round of cases did not just focus on Direct Disclosure, but also on broad exploration of several case themes: creative vs. malicious use, transparency via direct and indirect disclosure, and consent.

   We want to leave room for respondents to highlight any *other* areas of the Framework that can be deepened or improved upon to ensure its viability in a rapidly changing synthetic media ecosystem (related to the case themes above, and moving beyond the Direct Disclosure focus of this case template).

2. Has the Framework improved any processes, procedures, or policies at your organization, or your organization's observations of those Building, Creating, and/or Distributing synthetic media?