Artificial Intelligence (AI) for Sexual Assault Service Providers:

ETHICS & BEST PRACTICES

There are many considerations when making decisions about artificial intelligence (AI). The conversation around the ethical use of AI is often discussed in all-or-nothing terms, but there is more nuance to it than that. Every organization and individual must make their own choices about how they use AI, based on their values and circumstances. This supplemental resource (please also see WCASA's Introduction to AI resource) is intended to give you an overview of some of the ethical concerns related to AI, and related options for modifying your AI use.

CONCERNS & LIMITATIONS OF AI

There are some important ethical concerns to address related to AI use. This section highlights five primary concerns: accuracy, bias, environmental impact, plagiarism, and mental wellness. While some concerns are largely outside of the user's control, a better understanding of AI technology and the concerns around it can help you to use it more ethically and efficiently.

Accuracy

When AI Gets It Wrong: Addressing AI Hallucinations and Bias | MIT Sloan Teaching & Learning Technologies

AI tools like ChatGPT are trained on large datasets – collections of text from the internet, books, and other sources. They do not actually "understand" meaning or verify information as factual. Instead, AI generates responses by predicting what words are most likely to appear together based on patterns in its training data. This allows AI to produce fluent, natural-sounding text, but it also means the information it provides is not always reliable. AI may present misleading or entirely false information – a problem often called "hallucination."

Key Points to Remember:

- AI generates text, not necessarily factual content.
- AI may present misinformation in a confident tone.
- Hallucinations are more common in technical or specialized tasks, such as advanced math and legal analysis.
- AI-generated text should be fact-checked with credible, up-to-date sources.

Bias

Racism and AI: "Bias from the past leads to bias in the future" | OHCHR

Closely related to concerns around accuracy, AI is prone to biases. AI datasets include text that reflects inaccuracies, stereotypes, and systemic inequities. This means AI outputs can reinforce bias and erase opinions that do not align with the dominant narrative. There is even more concern

about this bias causing harm to historically marginalized communities as AI-support technology expands in schools, workplaces, policing, and other systems.

Key Points:

- All AI-generated content must be reviewed for stereotypes and/or biases.
- AI summaries tend to prioritize continuity by default, flattening nuance and erasing diverse or dissenting opinions.
- Culturally-specific or community-specific content should not be developed without authentic input. Like with any culturally-specific outreach, someone from the community should be generating the ideas and reviewing the content. Additionally, AI should not be used for translation or interpretation.
- People think of AI and technology as objective and neutral; this leads them to challenge its results less than they might human decisions.

Environmental

We did the math on AI's energy footprint. Here's the story you haven't heard. | MIT Technology Review

AI has significant environmental costs, including high electricity and water consumption. The amount of resources used depends on many things like the AI model, type of prompt, data center location, and time of day. Addressing the environmental costs of AI requires a more transparent and accountable technology industry, but individuals can still make choices about how they use AI to reduce their impact.

Key Points:

- Every AI prompt consumes energy; longer and more complex prompts and tasks use more energy.
- Images and videos use more energy to create than text.
- Smaller, task-specific AI models are more energy-efficient than large general-purpose ones.
- Transparency from AI companies is crucial to better understand and address environmental impacts. Corporate practices have a much greater influence on the overall impact of AI than individual ones.

Plagiarism

Is Using AI Plagiarism? | ACE Blog

While AI use is not inherently plagiarism, it can lead to plagiarism – intentionally or accidentally. If a human does not provide the ideas and content behind the work, AI will fill in gaps with the most common responses, which may be plagiarized or resemble other works. Funders and institutions may have different policies on AI use, and violations may disqualify work.

Key Points:

- Always check funder or institution policies before using AI for proposals or other products.
- AI-generated text needs to be supported by citations, just like human-generated text does.
- The safest process to avoid plagiarism: begin with your own writing → use AI for editing support → finish with human review.

- AI is not reliable for vetting and citing sources. Even when real sources are provided, they may be outdated, of questionable quality, or not accurately quoted.
- AI-powered plagiarism detection tools may help you find and address potential plagiarized text; these tools are imperfect and should be fact-checked like any other AI platform.

Mental Wellness
Exploring the Dangers of AI in Mental Health Care | Stanford HAI

Much like other technology, AI can impact mental health. Reported cases have been rare and often involve people with pre-existing conditions. More research is needed, but experts warn that AI use may exacerbate and/or reinforce delusions or suicidal thoughts in individuals experiencing crises related to self-harm or serious mental health conditions like psychosis. While the most serious outcomes have been rare, practicing good mental wellness while using technology is recommended for all people, and the recommendations around mental health can help you avoid other AI concerns.

Key Points:
- AI is designed to mirror users' tone and beliefs, even if they are harmful. General-use AI platforms like ChatGPT are not designed to recognize signs of a mental health crisis.
- Overestimating AI's intelligence or treating it as an all-knowing authority may contribute to negative mental health outcomes.
- AI cannot replace human connection, collaboration, or emotional support. Teamwork with real people benefits both mental health and work quality.

TIPS FOR MORE ETHICAL AI USE

AI can be a helpful tool, but using it responsibly and efficiently requires some care and intention. The following tips are designed to help users reduce risks, avoid harm, and make thoughtful decisions about when and how to incorporate AI into their work.

1. **Limit use to essential and appropriate tasks**: Use AI where it adds clear efficiency or value and rely on human expertise where nuance and context are crucial. Consider other existing tools that may be better suited to certain tasks.
2. **Keep prompts clear and concise**: Shorter prompts use less energy and tend to produce more accurate results. Review prompts before entering them to be sure they are as clear and direct as possible. Extra words such as "please," use energy without adding necessary information.
3. **Educate yourself about bias and stereotypes:** In order to review for something, you must be knowledgeable about it. Not all biased comments are as obvious as a slur being used. Learning about systems of oppression, stereotypes, and biases can help you better recognize harmful content.
4. **Give AI the context it needs for the task:** Think of what information the AI may go searching its training data for and aim to fill in as many gaps as you can. This helps ensure your content is based on your own ideas, and can limit bias and misinformation.

5. **Be critical of AI products**: Fact-check AI outputs, review for bias, and ensure citations are accurate, if they are provided. Consider using a plagiarism detector.
6. **Increase your AI literacy:** A realistic understanding of what AI can and cannot do helps support better AI use. This can help users make choices that protect their mental wellness, reduce their environmental impact, and avoid biased or inaccurate information in AI products.
7. **Respect policies and ethics**: Always follow institutional, funder, or organizational guidelines on AI use.
8. **Center human expertise**: Ideas and decisions should come from humans. Involve people with lived or cultural expertise to avoid harm or misrepresentation.
9. **Be transparent**: Acknowledge when and how AI has been used in your work, as needed and appropriate for the use. Humans tend to respond negatively to AI-generated creative work, so while this may be required in a grant application, it is not always advised or necessary for something like social media posts.
10. **Avoid over-reliance** – AI should support human work, not replace critical thinking, decision-making, accountability, or human connection. Monitor how AI affects your mental health, relationships, and time.

QUESTIONS TO GUIDE ETHICAL AI USE

There are many things to consider when developing an AI policy or deciding if AI is appropriate for a given task. Every agency has its own set of values and circumstances, and AI policy should reflect them. This list of questions is not intended to be a checklist every agency uses to review AI use, but rather for you to consider what parameters you might want to have in place for your organizational or individual AI use.

Accuracy and Verification

- Does the task require AI to provide factually accurate information?
- Can you verify the information you are asking AI to provide with reputable sources?
- Am I prepared to thoroughly review the AI's output for accuracy and bias? What might that review look like? Does it save time to use AI, based on what review might look like?

Bias and Representation

- Does the task involve culturally-specific or community-specific knowledge? Do you have the lived experience to verify that knowledge?
- Is nuance, context, or conflicting opinions important for this task? (ex. summarizing feedback from survivors)
- Is this a context (such as hiring, evaluations, or employee reviews) where bias in AI could be especially harmful? What are the potential consequences of bias in this context?
- Am I relying on AI for decision-making or judgment?

Plagiarism and Ethics

- Am I following all relevant funder, institutional, or organizational policies on AI use?

- Are the ideas and facts behind this product human-generated?
- Do I understand how to check for and avoid plagiarism in this task?
- Do I have a plan to locate and cite reliable references?

Environmental Responsibility

- Is AI the most energy-efficient tool for this task, or would a simpler method work as well or better?
- Is this task essential, or am I using AI for my own entertainment or curiosity?
- Can I make my prompts more concise to reduce resource consumption?
- Am I using the smallest, most appropriate model available rather than a large general-purpose one?

Mental Health and Wellbeing

- Does this task expect AI to act as an authority on personal, philosophical, or emotional issues?
- Am I maintaining healthy boundaries in my use of AI – balancing it with human connection, collaboration, and support?
- Could this task reinforce harmful beliefs, delusions, or emotional distress if handled by AI?

While not experts in AI, WCASA staff are actively engaging in AI use in our work. In fact, we used it to help us develop this resource.

If you need support, please reach out: wcasa@wcasa.org