

# Diffusion Models

# Ethics Case Studies

## AI and Art

The advent of AI art, powered by advanced generative models, has transformed the landscape of artistic creation and consumption. AI can generate artwork that is not only visually stunning but also pushes the boundaries of traditional art forms. However, this new frontier also raises significant ethical questions about authorship, originality, and the potential impact on the art community. This case study examines the ethical challenges associated with AI-generated art and explores the implications for artists, consumers, and society at large.

|  |
| --- |
| **Discussion Questions**   * What are the implications of AI art for traditional artists? How can the art community support human artists in an era of increasing AI involvement? * Who owns the rights to AI-generated art—the creator of the algorithm, the user who input the prompts, or the AI itself? * Should there be a distinction between AI-generated art and human-created art in competitions and exhibitions? * How might the rise of AI art change our understanding of creativity and artistic expression? Can AI introduce new styles, techniques, and forms of art. |

## Safety and generative models

Closed-source models like DALL-E, Midjourney, and others have implemented stringent safety measures to mitigate the risks of misuse. These models often come with built-in content filters to prevent the generation of harmful, inappropriate, or offensive material. For instance, DALL-E restricts the creation of images depicting violence, adult content, and other sensitive subjects. These guardrails are designed to ensure that the generated content adheres to ethical guidelines and complies with societal norms.

Open-source models like Stable Diffusion have faced significant scrutiny due to their inherent accessibility, which can be both a strength and a potential risk. The release of Stable Diffusion 3 introduced extensive safety guardrails to address ethical concerns. These measures include stricter content filtering and mechanisms to limit the generation of harmful content. Some community users argue that these safety measures undermine the principles of open-source and hinder legitimate creative freedom.

|  |
| --- |
| **Discussion Questions**   * How can open-source diffusion models balance the need for creative freedom with the responsibility to prevent misuse? * Are the current safety measures in models like Stable Diffusion 3 sufficient or too restrictive? * In what ways can safety guardrails in diffusion models hinder or promote innovation? Should there be different standards for closed-source versus open-source models? * How should the community be involved in defining and implementing ethical guidelines for open-source diffusion models? What are the potential benefits and drawbacks of community-driven content moderation? * Compare the ethical responsibilities of companies managing closed-source diffusion models with those developing open-source models. Should the expectations and standards for ethical practices differ between these two types of models? |

### References

<https://openai.com/index/dall-e-3/>

<https://openai.com/index/reducing-bias-and-improving-safety-in-dall-e-2/>

<https://openai.com/index/child-safety-adopting-sbd-principles/>

## GenAI in Academia

The rise of image generation models has led to speculation about how models like DALL-E are being used in academia. In a notable incident, several papers in biomedical research were retracted because they included AI-generated images that were not only scientifically inaccurate but also visually absurd. These images had passed through peer review undetected, raising serious concerns about the robustness of the review process. Although many journals have policies around *language* models, explicit guidance regarding image generation has lagged behind.

A diagram of a cell

Description automatically generated

|  |
| --- |
| **Discussion Questions**   * What scientific research tasks (if any) are acceptable to complete with image generation models? * Why are some tasks acceptable while others are not? Does this differ across disciplines? * Can we finetune models specially aimed at producing academic figures? * Should the model be credited as an author? * Should journals require a declaration from authors regarding the use of AI-generated content in their submissions? How might such a policy be effectively enforced? * How important is transparency in the use of generative AI for maintaining the credibility of scientific research? |

### References

https://venturebeat.com/ai/science-journal-retracts-peer-reviewed-article-containing-ai-generated-nonsensical-images/