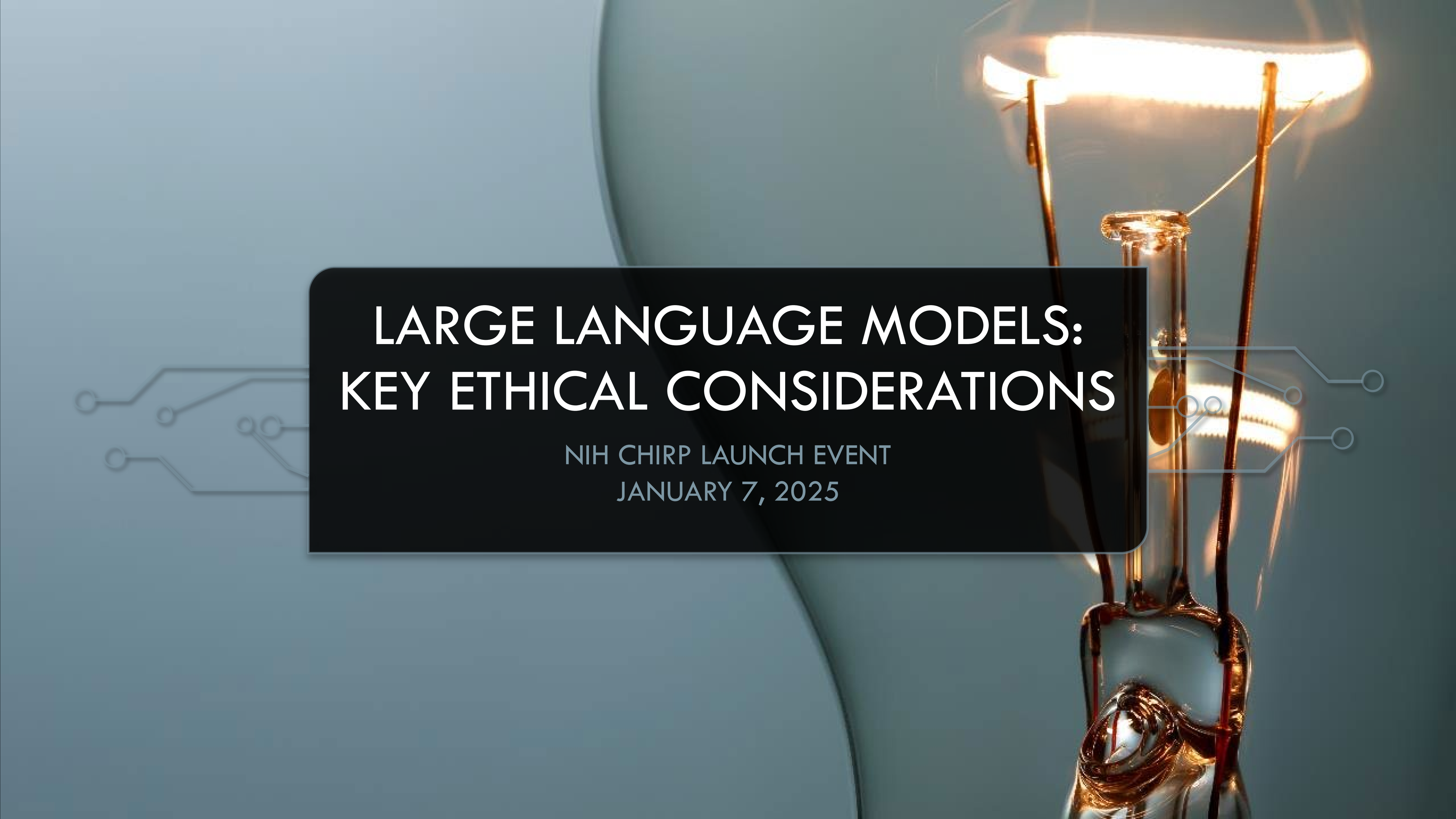
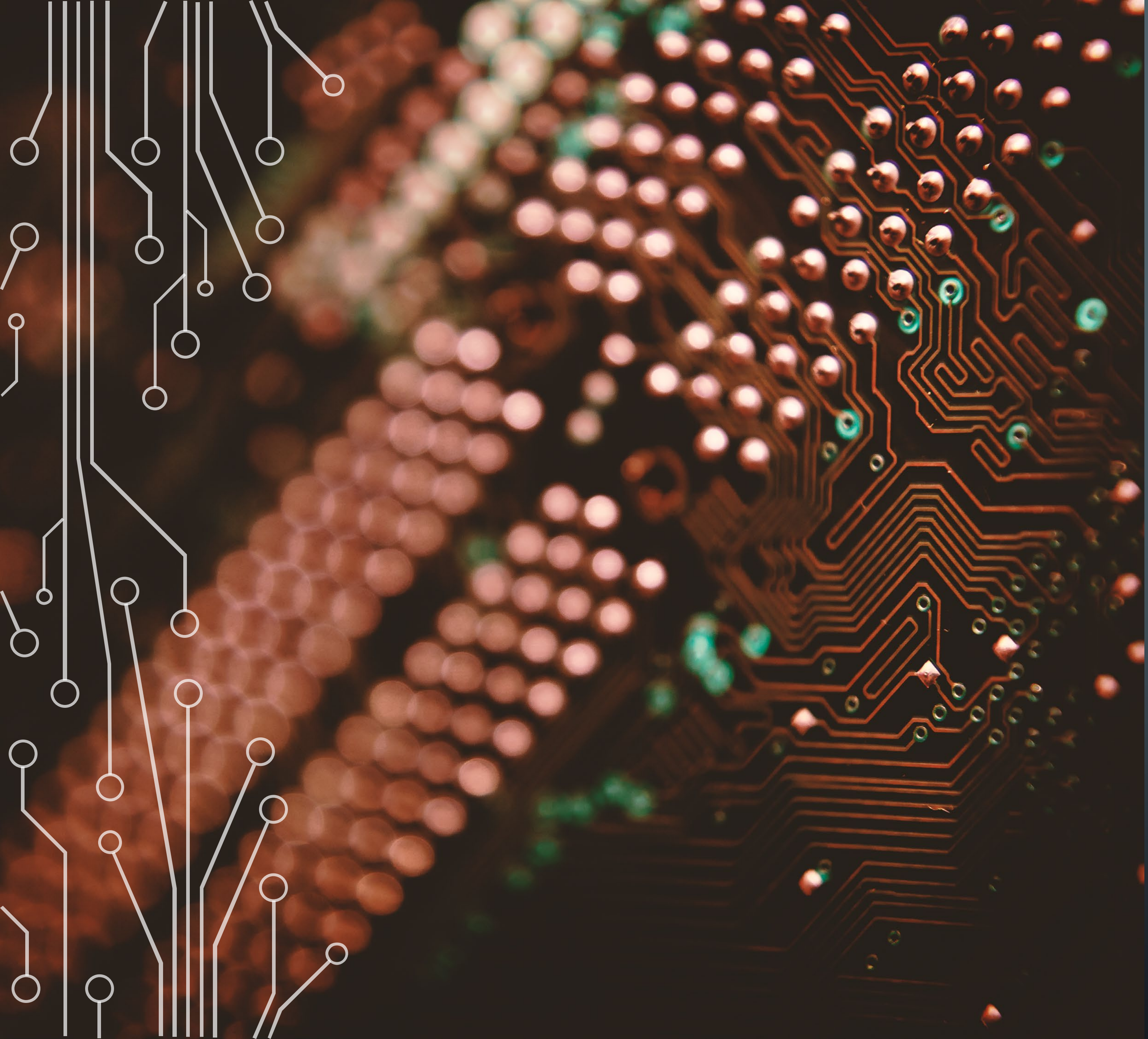


LARGE LANGUAGE MODELS: KEY ETHICAL CONSIDERATIONS

NIH CHIRP LAUNCH EVENT
JANUARY 7, 2025



TOPICS

- What’s *Responsible Use* of LLMs?
- Key Ethical Considerations
- AI Resources
- Questions

WHAT'S *RESPONSIBLE USE?*

- Utilizing Large Language Models (LLMs) with awareness of their limitations, potential bias, and ethical considerations, ensuring they are used as tools to enhance human capabilities rather than replace human judgement, and always verifying the accuracy and reliability of the information they generate before relying on it fully.

SIX (6) KEY ETHICAL CONSIDERATIONS

- Bias & Fairness
- Privacy & Security
- Transparency & Accountability
- Misinformation & Harmful Content
- Intellectual Property & Plagiarism
- Autonomy & Human Agency

BIAS & FAIRNESS

- **Issue:** LLMs can perpetuate and amplify biases present in the data they are trained on, such as gender, racial, and cultural biases.
- **Ethical Concern:** These biases can result in discriminatory outcomes or reinforce harmful stereotypes.
- **Mitigation:** Ongoing research is needed to reduce bias in LLMs, including better data curation, fairness audits, and transparency in model design.

PRIVACY & DATA SECURITY

- **Issue:** LLMs are trained on vast amounts of data, some of which may contain sensitive personal information.
- **Ethical Concern:** Using LLMs can inadvertently compromise privacy by generating or recalling sensitive data.
- **Mitigation:** Implementing strong data anonymization techniques, limiting access to personal data, and ensuring robust privacy safeguards in the development and deployment process.

TRANSPARENCY & ACCOUNTABILITY

- **Issue:** LLMs often operate as "black boxes," where users may not fully understand how decisions or outputs are made.
- **Ethical Concern:** Lack of transparency can make it difficult to hold LLMs accountable for harmful or incorrect outputs.
- **Mitigation:** Efforts to improve model interpretability and accountability mechanisms, such as clear documentation of model behavior and decision-making processes.

MISINFORMATION & HARMFUL CONTENT

- **Issue:** LLMs have the potential to generate misleading or harmful information, including fake news, harmful medical advice, or incitement to violence.
- **Ethical Concern:** The spread of misinformation can lead to societal harm and undermine trust in information.
- **Mitigation:** Incorporating content moderation systems, robust fact-checking processes, and setting ethical guidelines for model use.

INTELLECTUAL PROPERTY & PLAGIARISM

- **Issue:** LLMs can generate content that may unintentionally resemble copyrighted material, raising concerns about plagiarism or intellectual property theft.
- **Ethical Concern:** LLMs must be used in ways that respect intellectual property laws and avoid unintentional plagiarism.
- **Mitigation:** Encourage the use of proper citations and licenses, as well as encouraging users to review the content before publication.

AUTONOMY & HUMAN AGENCY

- **Issue:** Over-reliance on LLMs in decision-making processes may undermine human autonomy, leading individuals to blindly trust automated outputs.
- **Ethical Concern:** Automation should complement human decision-making, not replace it, to preserve personal responsibility and agency.
- **Mitigation:** Establish boundaries on when LLMs should be used, ensuring humans remain in control of important decisions.

AI RESOURCES

- **NIH Artificial Intelligence (AI) Cybersecurity Guidance**

  https://wiki.ocio.nih.gov/index.php/NIH_Artificial_Intelligence_(AI)_Cybersecurity_Guidance
- **HHS Policy for Securing Artificial Intelligence (AI) Technology**
  https://intranet.hhs.gov/policy/hhs-policy-securing-artificial-intelligence-technology
- **OPM Responsible Use of Generative Artificial Intelligence for the Federal Workforce**
  https://www.opm.gov/data/resources/ai-guidance/
- **IBM SkillsBuild online course entitled Artificial Intelligence Fundamentals**
  https://skillsbuild.org/adult-learners/explore-learning/artificial-intelligence
  - This free, 10-hour course consists of 6 sub-courses and, if completed, results in a credential.
- **Introduction to the AI Guide for Government, GSA - IT Modernization Centers of Excellence**
  https://coe.gsa.gov/coe/ai-guide-for-government/introduction/index.html

CONTACT INFORMATION

NIH Privacy Program
Website: https://oma.od.nih.gov/DMS/Pages/Privacy-Program.aspx
Email: privacy@mail.nih.gov

List of NIH ICO Privacy Coordinators -
https://oma.od.nih.gov/DMS/Pages/Privacy-Program-Privacy-Coordinators.aspx

NIH Information Security Program
Website: https://ocio.nih.gov/
Email: nihinfosec@nih.gov

List of NIH ICO Information Systems Security Officers
https://ocio.nih.gov/nih/portal/information-security-program/nih-information-security-policies-standards/nih-information

Questions?

Presentation content generated by:
+ Google AI Overview
* ChatGPT