Tech Ethics LAB

TECH ETHICS FORUM

2024 CFP Projects:

The Ethics of Large-Scale Models

A forum showcasing the research and outcomes of the 2024 CFP projects funded by the Notre Dame–IBM Technology Ethics Lab

JANUARY 22–23, 2025 ▪ UNIVERSITY OF NOTRE DAME

IBM

UNIVERSITY OF NOTRE DAME | ETHICS AND THE COMMON GOOD

INSTITUTE FOR
Ethics and the
Common Good

The **Notre Dame–IBM Technology Ethics Lab**, a critical component of the Institute for Ethics and the Common Good and the Notre Dame Ethics Initiative, promotes interdisciplinary research and policy leadership in technology ethics and is supported by a $20 million investment from IBM.

IBM + UNIVERSITY OF NOTRE DAME

The plans for the Notre Dame-IBM Technology Ethics Lab were formed when it became clear that emerging technologies were beginning to face critical and complex ethical challenges. These technologies are making game-changing innovations at speeds never imagined, but they risk quickly losing the trust of the very society they intend to improve. Industry must prove to the world that we can be responsible with these incredible innovations.

By combining IBM's deep technical expertise with Notre Dame's strength in philosophy and ethics, we hoped to raise awareness and help other organizations address ethical issues. IBM made a 10-year, $20 million commitment to the Tech Ethics Lab to create a holistic approach to tech ethics and practical, research-based models for the ethical development and deployment of these technologies.

In a short amount of time, the Tech Ethics Lab has moved to the forefront of the global conversation on how to build ethics into the early stages of design and development and continue throughout the entire lifecycle of the technology. Furthermore, as the technologies evolve, so must the ethical lenses applied to those technologies.

Through IBM's partnership with Notre Dame, we are bringing together leading academic, business, community, and government leaders to ask the tough questions and to champion responsible technology that is human-centric. IBM is proud to team with Notre Dame to bridge the concept of putting principles into practice with concrete approaches and models that organizations can leverage across the globe.

Learn more about the Lab and its work at **techethicslab.nd.edu**.

LAB ADMINISTRATION

Sara Berger, IBM Lab Director and Senior Research Scientist, Responsible and Inclusive Technology

Adam Kronk, Notre Dame, Director of Research and External Engagement, Institute for Ethics and the Common Good

Jungmin Lee, IBM Associate Director of Lab Operations

Yong Suk Lee, Notre Dame, Program Chair for Technology Ethics, Institute for Ethics and the Common Good

Nuno Moniz, Notre Dame, Notre Dame-IBM Technology Ethics Lab and Associate Research Professor, Lucy Family Institute for Data & Society

Catherine M. Quinlan, IBM, Legal M&A Integration Executivae, IBM Technology

Meghan Sullivan, Notre Dame, Director, Ethics Initiative and Institute for Ethics and the Common Good; Wilsey Family College Professor of Philosophy

STEERING COMMITTEE

Nick Fehring, IBM, Controller

David Go, Notre Dame, Vice President and Associate Provost for Academic Strategy, Viola D. Hank Professor of Aerospace and Mechanical Engineering

Christina Montgomery, IBM, Vice President, Chief Privacy & Trust Officer

Jeffrey F. Rhoads, Notre Dame, Vice President for Research, Professor of Aerospace and Mechanical Engineering

Meghan Sullivan, Notre Dame, Director, Ethics Initiative and Institute for Ethics and the Common Good; Wilsey Family College Professor of Philosophy

Jeffrey J. Welser, IBM, Chief Operating Officer, IBM Research; Vice President, Exploratory Science and University Collaborations

TECH ETHICS FORUM ▪ NOTRE DAME–IBM TECHNOLOGY ETHICS LAB

January 22–23, 2025 | University of Notre Dame

SCHEDULE AT A GLANCE

All conference sessions will be held in McKenna Hall, Rooms 215/216.

Wednesday, January 22, 2025

Wednesday morning sessions are for CFP recipients and invited guests only; Wednesday afternoon and Thursday sessions are open to the campus community.

9:00 a.m.	Welcome to Invited Guests: Adam Kronk and Nuno Moniz, University of Notre Dame
9:15 a.m.	Lightning Talks
10:30 a.m.	Coffee Break (Lobby)
10:45 a.m.	Keynote: Nitesh Chawla, University of Notre Dame
11:30 a.m.	Administration Session: Natalia Tipton Hernandez, University of Notre Dame
12:00 p.m.	Lunch (Rooms 205/206/207)
2:00 p.m.	Introduction to Forum: Adam Kronk and Nuno Moniz, University of Notre Dame
2:15 p.m.	**Panel #1: Can We Rely on AI?** This panel addresses trust and dependability in artificial intelligence systems. How do we measure and validate AI system reliability? What technical and ethical frameworks should guide our trust in these systems? The panel will examine how we can evaluate AI systems' trustworthiness through a multidisciplinary lens while understanding their limitations and developing robust frameworks for assessing their dependability in critical applications across various domains. *Keynote: Hoda Heidari, Carnegie Mellon University; Josep Domingo-Ferrer, Universitat Rovira i Virgili; Toby Jia-Jun Li, University of Notre Dame; Katherine Elkins, Kenyon College*
3:45 p.m.	Coffee Break (Lobby)
4:00 p.m.	**Panel #2: How Do We Build Responsible AI?** This session delves into the practical challenges and methodological approaches to developing ethical AI systems. What concrete steps can developers take to embed ethical considerations into AI systems, and how do we verify that these systems align with human values? The session explores strategies for responsible innovation, methodologies for identifying and mitigating bias, and best practices for ensuring AI systems serve societal needs throughout their development lifecycle. *Keynote: Ricardo Baeza-Yates, Northeastern University Institute for Experiential AI; Katherine Walden, University of Notre Dame; Franziska Poszler, Technical University of Munich; Angélica García Martínez, University of Notre Dame*
5:30 p.m.	Transition to Reception: Scott Graham, University of Notre Dame
5:45 p.m.	Reception and Dinner for Invited Guests (Seven on 9, Corbett Family Hall, Door 3; map on inside back cover of program)

Thursday, January 23, 2025

9:00 a.m.	Review of First Day: Adam Kronk and Nuno Moniz, University of Notre Dame
9:15 a.m.	**Panel #3: Labor in the Age of AI** This session examines the transformative impact of artificial intelligence on the workforce and labor markets. How is AI reshaping the nature of work, and what new skills will workers need to thrive in an AI-augmented workplace? The discussion encompasses both immediate challenges and long-term implications, including workforce adaptation, economic displacement, labor rights, and strategies for ensuring equitable distribution of AI's benefits in the workplace. *Keynote: Yong Suk Lee, University of Notre Dame; Lisa van der Werff, Dublin College; Avigail Ferdman, Technion Israel Institute of Technology*
10:45 a.m.	Coffee Break (Lobby)
11:00 a.m.	**Panel #4: Education in the Age of AI** This panel explores how artificial intelligence is changing teaching and learning environments. How can AI enhance rather than replace human-centered learning, and what role should it play in assessment and personalized education? The session examines AI's role in transforming educational practices, implications for pedagogy, educational equity, and the balance between technological innovation and meaningful learning experiences. *Keynote: Ron Metoyer, University of Notre Dame; Felix Kayode Olakulehin, National Open University of Nigeria; Ranjit Singh, Data & Society Research Institute; Alison Cheng, University of Notre Dame*
12:30 p.m.	Lunch (Rooms 205/206/207)
1:45 p.m.	**Panel #5: Memory, History, and AI** This session investigates the complex relationship between artificial intelligence and our understanding of the past. How does AI technology influence our interpretation and preservation of historical narratives, and what are the ethical implications of using AI to process and analyze historical data? The panel explores AI's impact on historic preservation, interpretation, and documentation and its role in shaping collective memory and cultural heritage. *Keynote: Jasna Ćurković Nimac, Catholic University of Croatia; Emillie de Keulenaar, University of Groningen; Luis Gabriel Moreno Sandoval, Pontificia Universidad Javeriana; Jamie Kelly, Vassar College*
3:15 p.m.	Coffee Break (Lobby)
3:30 p.m.	**Panel #6: Policy, Governance, and Ways Forward** This panel addresses the crucial challenge of developing effective AI oversight and regulation frameworks. What governance structures can effectively guide AI development while fostering innovation, and how do we implement meaningful oversight that protects public interests? This forward-looking session examines approaches to governing AI development and deployment, focusing on practical policy solutions and regulatory mechanisms that promote responsible AI advancement. *Keynote: Marianna B. Ganapini, Union College; Catherine Botha, University of Johannesburg; Georgina Curto Rex, University of Notre Dame; Paolo G. Carozza, University of Notre Dame*
5:00 p.m.	Closing Remarks: Adam Kronk and Nuno Moniz, University of Notre Dame
5:15 p.m.	Reception (Lobby)

2024 CFP PROJECTS

The Lab's 2024 Call for Proposals delved into the moral dilemmas of large-scale models, their potential impacts, and efforts to navigate these challenges responsibly. A total of $942,117 was awarded to 17 projects. Eleven projects were completed by December 31, 2024; six were continued through June 30, 2025.

Large-scale models are a subset of artificial intelligence systems designed to perform a range of tasks by learning from vast amounts of data. Unlike traditional AI systems, LSMs are pre-trained on extensive datasets and can be fine-tuned to perform various functions. The versatility of LSMs has led to widespread adoption across different sectors, from education and healthcare to entertainment and communication. However, rapid adoption comes with various ethical challenges, including bias, misinformation, privacy concerns, and intellectual property rights. Addressing these issues is pivotal, as the future of AI depends on the responsible and transparent use of these models.

The ethics of LSMs revolve around several core issues, ranging from societal biases to the spread of misinformation. Each of these concerns needs to be examined closely to mitigate potential risks and ensure the responsible deployment of these technologies.

CRITICAL THINKING IMPACT

There is rising concern that over-reliance on AI-generated content could weaken analytical reasoning and human expertise. In areas like education and healthcare, excessive dependence may lead to automation bias, where professionals trust AI outputs without thoroughly questioning their accuracy.

PRIVACY CONCERNS

LSMs rely on vast datasets that may contain personal information, raising privacy concerns. Without proper safeguards, sensitive data can be exposed or misused. Additionally, their complexity often obscures how outputs are generated, undermining accountability and transparency.

INTELLECTUAL PROPERTY ISSUES

The rise of LSMs has ignited debates over intellectual property and human rights in AI-generated content. Traditional IP frameworks, designed for human creativity, struggle to address AI's unique challenges.

BIAS AND DISCRIMINATION

Because these models are trained on vast datasets that often include biased or unbalanced information, they can reflect and reinforce discriminatory attitudes toward race, gender, or ethnicity. This issue is especially concerning in hiring, law enforcement, and content creation, where biased outcomes can result in significant consequences.

MISINFORMATION AND FAKE NEWS

With their ability to produce human-like text, LSMs can be exploited to create fake news, misleading articles, and even deepfake content. The consequences are far-reaching, as misinformation can undermine public trust, manipulate political outcomes, and cause widespread confusion.

FUNDED 2024 CFP PROJECTS

1) Conflict Moderation: Implementing Bridge-Building Design in LLM Models Principal Investigator: Emillie de Keulenaar (University of Groningen); Notre Dame Collaborator: Lisa Schirch, Keough School of Global Affairs

2) Contextualizing AI Ethics in Higher Education: Comparing the Ethical Issues Raised by Large-Scale Models in Higher Education Across Countries and Subject Domains Principal Investigator: Wayne Holmes, (Institut "Jožef Stefan"); Caroline Pelletier (Institut "Jožef Stefan"); Notre Dame Collaborator: Ying (Alison) Cheng, Psychology

3) Cultural Context-Aware Question-Answering Systems: An Application to the Colombian Truth Commission Documents Principal Investigator: Luis Gabriel Moreno Sandoval (Pontificia Universidad Javeriana); Maria Prada Ramirez and Anna Sokol (Pontificia Universidad Javeriana); Notre Dame Collaborator: Matthew Sisk, Lucy Family Institute for Data & Society

4) Engaging End Users in Surfacing Harmful Algorithmic Behaviors in Large-Scale AI Models Principal Investigator: Wesley Hanwen Deng (Carnegie Mellon University); Motahhare Eslami, Ken Holstein, and Jason Hong (Carnegie Mellon University); Notre Dame Collaborator: Toby Jia-Jun Li, Computer Science and Engineering

5) Ethical Deployment of Generative AI Systems in the Public Sector: A Practitioner's Playbook Principal Investigator: Titiksha Vashist (The Pranava Institute); Dhanyashri Kamalakkannan and Shyam Krishnakumar (The Pranava Institute); Notre Dame Collaborator: Georgina Curto Rex, Lucy Family Institute for Data & Society

6) Ethical LLM-based Approach to Improve Early Childhood Development in Children with Cancer in LMICs Principal Investigator: Horacio Márquez-González (Hospital Infantil de México Federico Gómez); Notre Dame Collaborators: Nitesh Chawla and Angélica García Martínez, Lucy Family Institute for Data & Society

7) Generative AI and the Social Value of Artifacts: The Case for Saving Photo Morgues Principal Investigator: Jamie Kelly (Vassar College); Kafui Attoh (CUNY School of Labor and Urban Studies); Notre Dame Collaborator: Don Brower, Center for Research Computing

8) How LLMs Modulate our Collective Memory and its Ethical Implications Principal Investigator: Jasna Ćurković Nimac (Catholic University of Croatia); Ivana Brstilo Lovríc (Catholic University of Croatia); Notre Dame Collaborator: Nuno Moniz, Notre Dame-IBM Technology Ethics Lab, Lucy Family Institute for Data & Society

9) How Well Can GenAI Predict Human Behavior? Auditing State-of-the-Art Large Language Models for Fairness, Accuracy, Transparency, and Explainability (FATE) Principal Investigator: Katherine Elkins (Kenyon College); Jon Chun (Kenyon College); Notre Dame Collaborator: Yong Suk Lee, Keough School of Global Affairs

10) Impact of Generative Artificial Intelligence—ChatGPT—on Higher Education in the Global South: Ethics and Sustainability Principal Investigator: Helen Titilola Olojede (National Open University of Nigeria); Felix Kayode Olakulehin (National Open University of Nigeria); Notre Dame Collaborator: Nitesh Chawla, Lucy Family Institute for Data & Society

11) LLMs and a Well-Rounded Approach to Human Flourishing Principal Investigator: Avigail Ferdman (Technion-Israel Institute of Technology); Leora Sung (Technion-Israel Institute of Technology); Notre Dame Collaborator: Don Howard, Philosophy

12) Mitigating Ethical Risks in Large Language Models through Localized Unlearning Principal Investigator: Josep Domingo-Ferrer (Universitat Rovira i Virgili); Alberto Blanco-Justicia, Najeeb Jebreel, and David Sánchez (Universitat Rovira i Virgili); Notre Dame Collaborator: Nuno Moniz, Notre Dame-IBM Technology Ethics Lab and Lucy Family Institute for Data & Society

13) Research-Based Theater: An Innovative Method for Communicating and Co-Shaping AI Ethics Research & Development Principal Investigator: Franziska Poszler (Technical University of Munich); Anastasia Aritzi and Christoph Lütge (Technical University of Munich); Notre Dame Collaborator: Carys Kresny, Film, Television, and Theatre

14) Seeing the World through LLM-Colored Glasses—Detecting Biases and Deficiencies in Language Model Presentation of Underrepresented Topics Principal Investigator: Ricardo Baeza-Yates (Northeastern University Institute for Experiential AI); Muhammad Ali, Shiran Dudy, Resmi Ramachandranpillai, and Thulasi Tholeti (Northeastern University Institute for Experiential AI); Notre Dame Collaborator: Toby Jia-Jun Li, Computer Science

15) Technology Transfer and Culture in Africa: Large Scale Models in Focus Principal Investigator: Catherine Botha (University of Johannesburg); Franklyn Echeweodor, Anthony Isong, and Edmund Ugar (University of Johannesburg); Notre Dame Collaborator: Georgina Curto Rex, Lucy Family Institute for Data & Society

16) The Ethics of Using Large-Scale Models: Investigating Literacy Interventions for Generative AI Principal Investigator: Emnet Tafesse (Data & Society Research Institute); Ranjit Singh, (Data & Society Research Institute); Notre Dame Collaborator: Karla Badillo-Urquiola, Computer Science and Engineering

17) The Influence of Virtual Avatar Race and Gender on Trust and Performance: Understanding How the Appearance of LLM-Enabled Avatars Influences Work in Virtual Reality Principal Investigator: Lisa van der Werff (Dublin City University); Theo Lynn (Dublin City University); Notre Dame Collaborator: Timothy Hubbard, Management & Organization

2025 CFP PROJECTS

The Lab is proud to announce the recipients of its 2025 Call for Proposals, which focuses on how to design effective solutions for safe and ethical human-AI collaboration in real-world settings. The objective is to foster developments that leverage AI to augment human tasks and ensure that these collaborations are ethical, inclusive, and beneficial to society at large. As AI systems become increasingly sophisticated, there is a pressing need to understand and optimize how they can work alongside humans in various domains and to anticipate potential medium- to long-term impacts of such applications in critical sectors. The Lab received many outstanding applications. Upon review, the Lab's selection committee awarded a total of $345,000 to six projects from the 99 applications received. The selected projects are listed below. Teams will have until December 31, 2025, to complete the projects. The reduction in the number of projects supported in the 2025 Call for Proposals as compared to 2024 signals a shift in the strategy of the Lab to focus on other additional collaborations, which we are confident will facilitate the greatest impact over the next five years.

Building AI Text Classifiers with Peacebuilders: A Human-AI Collaboration to Improve Conflict Analysis and Resolution Principal Investigator: Allan Cheboi (Build Up) ; Team Members: Julie Hawke, Lisa Schirch, Will O'Brien (University of Notre Dame); Notre Dame Collaborator: Lisa Schirch, Peace Studies

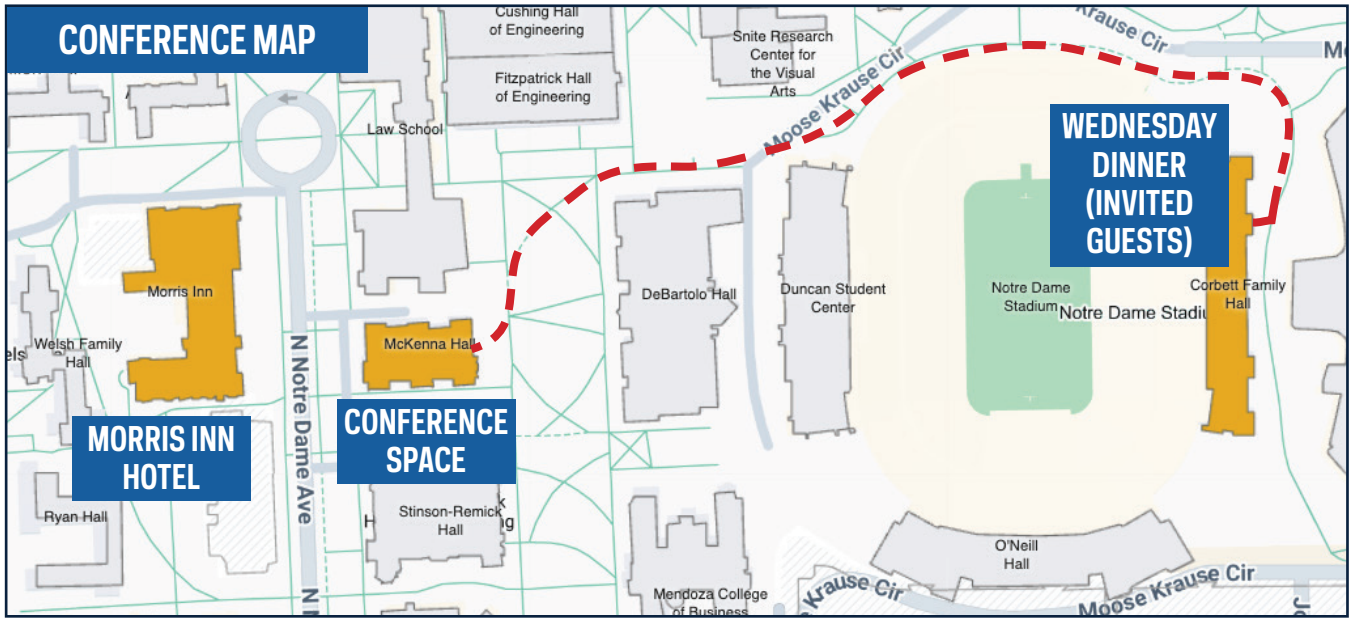
Digital Afterlives? Mourning, Memory, and Grief Tech Principal Investigators: Joseph Davis (University of Virginia) and Micah E. Lott (Boston College); Team Member: William Hasselberger (Catholic University of Portugal); Notre Dame Collaborator: Paul Scherz, Theology

Digital Moral Twins: From Bioethical Principles to AI Ethics and Back Again Principal Investigator: Jeffrey P. Bishop (Saint Louis University); Team Members: Emily Dumler-Winckler (Saint Louis University), Lydia Dugdale, MD (Columbia University Medical Center), Jason T. Eberl (Saint Louis University), S. Matthew Liao (New York University); Devan Stahl (Baylor University); Notre Dame Collaborator: Paul Scherz, Theology

(Digital) Companionship in the Digital Age: On Human-AI Relationships and the Ethical Landscape Surrounding Artificial Others Principal Investigators: Robert Clowes (NOVA University of Lisbon) and Kesavan Thanagopal (University of Notre Dame); Notre Dame Collaborator: Diego Gómez-Zará, Computer Science

Enhancing Human-AI Collaboration and Policy in Emergency Response: Ethical Deployment of AI-Enabled Drones Principal Investigator: Ricardo Morales (Brown University); Team Members: Kaitlin Harris (US Airforce/SAF/AQRE); Demetrius Hernandez (University of Notre Dame), Tristian Hernandez; Notre Dame Collaborator: Jane Cleland-Huang, Computer Science and Engineering

Image Descriptions Are Less Reliable Than They Appear: Support for Blind Users Assessing Capabilities of AI-Powered Access Technology Principal Investigator: Amy Pavel (University of Texas at Austin); Team Member: Meng Chen (University of Texas at Austin); Notre Dame Collaborator: Toby Li, Computer Science



https://techethicslab.nd.edu

UNIVERSITY OF NOTRE DAME | ETHICS AND THE COMMON GOOD

1124 Flanner Hall, Notre Dame, IN 46556 USA | (574) 631-1305 | ethics.nd.edu