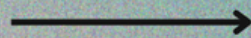


↓
2nd Sawyer Philosophy
Conference



11/08/2025

Ethics of Technology

Promises and Perils



2nd Sawyer Conference: Technology and Ethics: Promises and Perils

Schedule

"Please note that there will be parallel sessions in Morris Hall and in DTC CK 590 (5th Floor)."

Morris Hall

9:00 am Welcome, Elisabeth Hildt, Illinois Institute of Technology

9:15-10:45 am

Session 1: Machine Learning and Philosophy of Science

Chair: Kathryn Petrozzo

DAGnabbit: how Machine Learning is changing what *causation* means in science
Carlos Santana, University of Pennsylvania

Epistemic issues with scientific uses of synthetic data
Clarissa Busch, Illinois Institute of Technology

Committing speech acts without a common ground: Social epistemic risks of artificial conversation partners
Duygu Uygun-Tunç, University of Chicago

10:45 am - 11:00 am: Coffee Break

11:00-12:30 am

Session 2: Ethical Issues

Chair: Sydney Harvey

Human-Centered AI in Design: Ethical Opportunities and Challenges through the Lens of Design Thinking
Yun Dong, Illinois Institute of Technology

Multispecies Values in Urban Architecture: A Deweyan Account of Interactivity through Space and Place
Bjørn Ralf Kristensen, Illinois Institute of Technology

Poetically Man Dwells Upon the Earth
Caramia Axland, Colorado State University

12:30 pm - 1:30 pm: Lunch

1:30 pm-2:30 pm

Session 3: Medical Ethics & Tech

Chair: Clarissa Busch

Should an Algorithm Decide if I get Antibiotics? Algorithmic Decision-Makers in the Hospital and the Virtues of Acknowledged Dependence
Charles Freiberg, University of Central Florida

Promises & Perils of Neurotechnology in Clinical Psychiatry
Kathryn Petrozzo, Illinois Institute of Technology

2:30 - 2:45 pm: Coffee Break

2:45-4:15 pm

Session 4: Decolonizing AI

Chair: Bjørn Ralf Kristensen

Whose Knowledge Counts? Refusal, Embodiment, and the Ethics of Decolonizing AI
Gayatri Raman, Indiana University

Does AI need to be “decolonized”?
Tom Buller, Illinois State University

Maqāṣid al-Sharī‘a as a Framework for Ethical AI: The Perspective of Humanities on Technology and Human Dignity
Soleh Hasan Wahid, State Islamic University of Ponorogo, Indonesia

DTC CK 590 (5th Floor)

9:15-10:45 am

Session 1: Linguistics, Art, and LLMs

Chair: Elisabeth Hildt

"*Beyond Thinking*: excess and meaning in thought and its others"
Mark Walter, Aurora University

When Should We Believe LLM Outputs? A Testimonial Theory
James Gillard, University of Texas at Austin

AI Chatbots and the Paradox of Fiction: (How philosophy of fiction can shed light on AI-human relations)
Sydney Harvey, Illinois Institute of Technology

10:45 am - 11:00 am: Coffee Break

11:00-12:30 am

Session 2: Law & Jurisprudence

Chair: Katthryn Petrozzo

Against AI Jurisprudence: Large Language Models and the False Promises of Empirical Judging
Dasha Pruss, University of Illinois, Chicago

The Application of the Rule of Law: Technological Assistance and the Ontic/Ethic/Epistemic Paradigm
Dennis Kamalick, Chicago Police Force Training Academy

Why Copyright must remain Human-Centric: An Ethical Analysis of AI and Copyright
Eylül Erva Akin, Bocconi University, Milan, Italy

12:30 pm - 1:30 pm: Lunch

1:30 pm-2:30 pm

Session 3: Responsibility

Chair: Kelly Laas

Beyond Good and Evil: Thinking AI within the Human Condition
Jurgita Imbrasaite, University of Bonn

The Limits of AI Understanding and What they Mean for Users
Mike Zahorec, Florida State University

2:30 - 2:45 pm: Coffee Break

2:45-4:15 pm

Session 4: Agency

Chair: Kelly Laas

Agency Collapse in Engineering
Nick Treanor, University of Edinburgh

Why Machines cannot be Full Moral Agents
Dane Leigh Gogoshin, University of Wisconsin

Many thanks to the Sawyer Foundation for generously supporting this conference.

To connect to Wifi:

1. Select the wireless network IIT-Connect
2. Open any web browser
3. When prompted for username and password enter

Username: sawyier

Password: 25a8ec

4. Agree to the terms and conditions
5. Click continue

Abstracts and Biographies of Presenters

Morris Room

9:00 am Welcome, Elisabeth Hildt, Illinois Institute of Technology

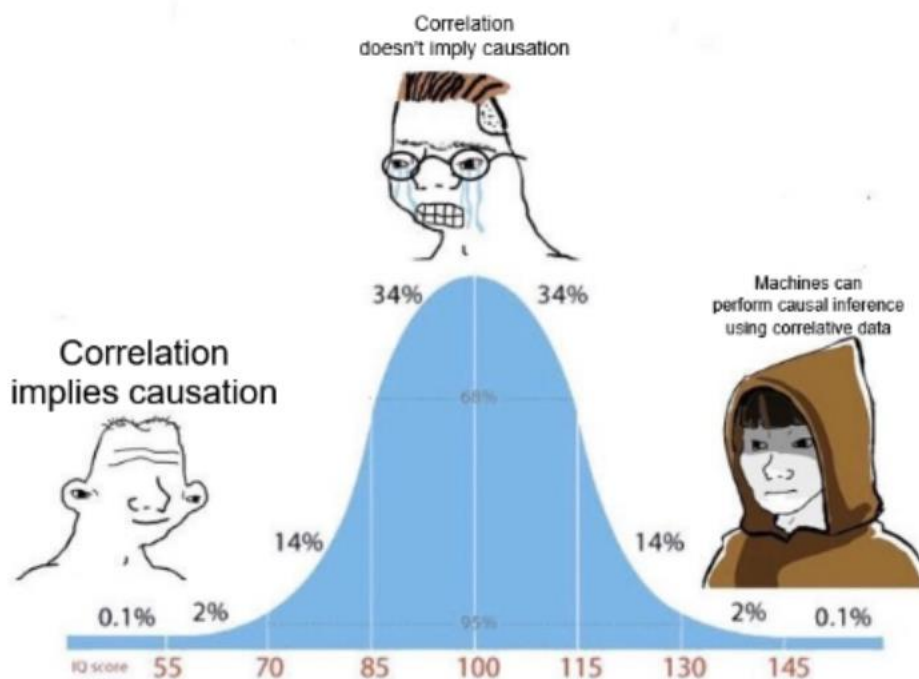
9:15-10:45 am

Session 1: Machine Learning and Philosophy of Science

Chair: Kathryn Petrozzo

DAGnabbit: how Machine Learning is changing what causation means in science
Carlos Santana, University of Pennsylvania

The only thing most of our students remember from their stats classes is that “correlation doesn’t imply causation.” For many questions in social sciences, epidemiology, and ecology, however, controlled experiments aren’t always practical or ethical, and we only have correlative data. Fortunately, new technologies are allowing for causal inference using observational data. For example, Structural Causal Modeling (SCM) (Pearl 2010) combines causal graphs with data-driven Machine Learning (ML) to infer causation from observational/correlative data.



Researchers are excited about the potential of SCM (Tennant et al. 2021; Arif and MacNeil 2022). It’s also a shiny new technology that risks unintended consequences. I argue that adopting machine learning methods for causal inference requires redefining causation in fields like social sciences, epidemiology, and ecology, and that this has significant consequences for the production of scientific understanding. We risk mutilitating scientific theory to fit the demands of our technology, rather than designing technology to facilitate better scientific theory.

Here’s the simple version of the argument:

The causal graphs required for automated (ML) causal inference are Directed Acyclic Graphs (DAGs). DAGs don't permit bi-directional causal interactions or feedback cycles. Established models in sciences like ecology and epidemiology are full of bi-directional causal interactions and feedback cycles. Thus, to adopt automated causal inference, these sciences have to replace established causal models with new DAG-compatible ones. We can do this, I'll argue, but need to do so carefully, or we'll sacrifice human understanding (kind of an important goal of scientific theory) to facilitate machine inference.

Carlos (they/he) thinks philosophy is great because it can help solve pressing problems. Right now that means things like addressing the loss of agricultural land in tropical regions, creating software to help local communities contextualize ecological science to their land and waters, and serving as an ambassador to help their university better integrate with the city of Philadelphia and Commonwealth of Pennsylvania. People often ask them, "But how is that philosophy?" Their response is that they're a philosopher in the same way that Indiana Jones is an archaeologist.

Epistemic issues with scientific uses of synthetic data

Clarissa Busch, Illinois Institute of Technology

The term "synthetic data", broadly speaking, refers to datasets in which the overall statistical patterns (or other relevant but non-identifying information) contained in the data are preserved, while the identities of the human data subjects from whom the data were initially collected are obscured. In recent years, synthetic data has gained popularity in machine learning, where it is increasingly being used to train large language models. This has led to a rise in critical engagement with the concept of synthetic data by scholars in critical data studies and philosophy (e.g., Susser et al. 2024, Lee et al. 2025). Since the central purpose of using synthetic data is often to preserve privacy, much of the critical literature has focused on the privacy-enhancing capacities of various synthetic data techniques and their implications for related ethical notions such as the privacy of collectives (Taylor et al. 2017).

This talk will explore the epistemic implications of the use of synthetic data. Specifically, I will identify various ways in which synthetic data is used in scientific contexts, differentiating between evidential and non-evidential uses. I will then explore the epistemic issues that arise from the various uses of synthetic data to support evidential claims.

Clarissa Busch (she/her) is an Assistant Professor of Philosophy at the Illinois Institute of Technology in Chicago. In her research, she brings a philosophy of social science lens to the climate change problem, as well as to ethical challenges related to AI technologies. Drawing on her dual background in philosophy and data science, she uses a diverse philosophical toolkit, including computational and statistical methods. Her current projects include papers on the philosophy of sustainability science, the moral psychology of climate action, and on social norms related to the adoption of AI tools.

Committing speech acts without a common ground: Social epistemic risks of artificial conversation partners

Duygu Uygun-Tunç, University of Chicago

This talk examines people's social-epistemic relationship to chatbots through the lens of our epistemic common ground, which is the normative structure of our social information environment on which testimonial knowledge depends. Human testimonial practices are sustained by shared expectations about sincerity, reliability, and accountability: when we take another's word, we rely not only on the content of their assertion but also on the background norms governing speech acts. However, people are often poor at discerning whether these expectations are indeed fulfilled. Unlike human interlocutors, chatbots participate in conversation without being bound by the same normative commitments, yet they often

appear to conform to them. This creates a distinctive epistemic vulnerability: users treat conversational outputs by chatbots as genuine testimonial acts, thereby extending trust in ways that place them in poor epistemic positions. I will argue that this misalignment is particularly dangerous in cases where chatbots offer unwarranted confirmations or validations. Reports of users being led to believe falsehoods—sometimes reassured by chatbots of the originality or soundness of mistaken ideas—illustrate how easily the illusion of common ground can mislead. The risk intensifies in sensitive contexts, such as when chatbots are used as ersatz psychotherapists, where misplaced epistemic trust may amplify vulnerability. By analyzing these dynamics, I aim to show how chatbots reshape the information environment and to highlight the urgent need for reflection on how to sustain our testimonial practices in this changing environment.

Duygu Uygun-Tunç is a Collegiate Assistant Professor and Harper-Schmidt Fellow at the University of Chicago, specializing in philosophy of science, epistemology, and philosophy of mind. Duygu received her PhD in Philosophy in 2020 from Heidelberg University and the University of Helsinki with a dissertation titled 'Communication and the Origins of Personhood'. Before her current appointment, she ran the research project 'Extended Scientific Virtue', funded by the European Union's Marie Skłodowska-Curie Actions, and worked on a Dutch Scientific Research Council project on evaluating interdisciplinarity in science. Duygu's current research focuses on social epistemology (of science and technology), research methodology in experimental social sciences, scientific integrity, social aspects of scientific knowledge, properties of epistemic systems, technologically extended forms of inquiry, as well as applied topics such as the replicability crisis and scientific reform in Psychology.

10:45 am - 11:00 am: Coffee Break

11:00-12:30 am

Session 2: Ethical Issues

Chair: Sydney Harvey

Human-Centered AI in Design: Ethical Opportunities and Challenges through the Lens of Design Thinking

Yun Dong, Illinois Institute of Technology

While Artificial Intelligence (AI) is rapidly transforming creative and design practices, much of the discourse has focused on technical advances rather than the ethical and human-centered implications. This presentation foregrounds a design perspective by using Design Thinking as a framework to examine AI's promises and perils.

The five stages of Design Thinking—Empathize, Define, Ideate, Prototype, and Test—are used to analyze how AI reshapes the design process. Opportunities include accelerating prototyping, expanding creative ideation, and democratizing access to design tools. Ethical challenges, however, emerge in issues of authorship, homogenization of outputs, algorithmic bias, and the erosion of human responsibility.

By situating AI within this human-centered framework, the presentation highlights how design methods can help articulate, navigate, and mitigate ethical tensions. These insights not only inform discussions of AI in design, but also resonate with related domains such as engineering education, where similar tensions between efficiency, integrity, and responsibility arise.

Yun Dong is an Assistant Teaching Professor jointly appointed in the Department of Information Technology and Management and the Department of Computer Science at the College of Computing, Illinois Institute of Technology. She holds a Ph.D. in Human-Computer Interaction, and her teaching and

research focus on user-centered design, human-computer interaction, and the applications of AI in game design and education. Her recent work explores how AI can enhance interactive learning environments, improve game-based design processes, and support innovative educational experiences that bridge design thinking with emerging technologies.

Poetically Man Dwells Upon the Earth

Caramia Axland, Colorado State University

Martin Heidegger postulates in “Question Concerning Technology” that technology enframes human beings and art is our saving power. Creating art is a process which operates as an expression of truth and the human condition. The advent of Artificial Intelligence (AI) generated “art” challenges this notion by calling into question the essence of art. Humans have control over the data sets given to the AI but little control over the end result, eliminating the actual process of creation, and negating any mastery on the part of the human user. AI includes computer-rendered artificial pictures which are referred to as “art”. Art, as Heidegger describes, is a saving power that will aid human beings in revealing truth and work against technological enframing. Reflection upon art, an innately human creation, assists in the reflection upon technology and the way in which technology enframes human life and experience, this is the saving power. The saving power has been threatened by the recent introduction and acceptance of AI-generated “art”. AI only generates results derived from pre-existing human creations, disregarding the emotive action necessary to creating art, and the human reality within the art itself. We ought not address the product of generative AI as “art”, for fear that “art” loses its power, rather as computer-rendered artificial pictures (C.R.A.P.). In the place of *aletheia*, C.R.A.P. present an approximation of the things once represented as themselves, distorted to fit an end. In this process, we lose human expression, art’s true essence.

Caramia Axland (they/them/theirs) is a first-year master's student at Colorado State University. Their areas of interest are incredibly broad phenomenology, philosophy of technology, ethics, aesthetics, philosophy of neurodiversity, and political philosophy.

Multispecies Values in Urban Architecture: A Deweyan Account of Interactivity through Space and Place

Bjørn Ralf Kristensen, Illinois Institute of Technology

Working through American pragmatist philosopher John Dewey’s naturalistic metaphysics, I explore agency and architecture, with a mind toward the place of nonhuman animals within urban environments. I begin with an account of Dewey’s concept of “interactivity,” which challenges ontological divisions between agent and environment. Central to this exploration is my claim: How humans come to understand other animals, and how we act and relate to one another is inscribed on us through everyday—and often otherwise mundane and unreflected upon—interactions with the social and material world. Next, I explore Dewey’s accounts of architecture as fluid, value-laden, and powerfully capable of reshaping experience. Another central claim of Dewey’s approach to architecture aligns with his consequentialist concept of “ends-in-view.” Dewey challenges moral-nonmoral dualism and puts forward an emphasis that there are ethical demands stemming from the reality that consequences of actions likely always exceed intentions. I use this grounding to approach questions of spatial agency and urban animals. What values are present through urban development for humans and other animals? What forms of agency? How are human perceptions of other animals influenced through particular decisions made in the urban environment? Many of the most influential factors in the lives of urban animals exceed intentional inclusion of them. To think through this approach, I consider one of the most common commensal urban species, both globally and in the context of Chicago, the pigeon (rock dove).

Bjørn is a Sawyer fellow with the Center for the Study of Ethics in the Professions at Illinois Institute of Technology and a PhD candidate in the Environmental Sciences, Studies, and Policy program at the University of Oregon, where he is also a full member of the Department of Philosophy. His research is focused on applied ethics in the contexts of animals and the environment, with particular emphasis on liminal animals, urban environments, waste, sanitation, and development. He works within the traditions of American philosophy and ecofeminism, with a focus on the work of John Dewey and Val Plumwood. His approach also pushes against hard disciplinary boundaries and often incorporates ethnographic field methods. In his dissertation project, he explores complex ethical questions arising from the lives of free-ranging street dogs in Cusco, Peru, which challenge norms within Western animal welfare frameworks. In the Fall semester at Illinois Tech, he is teaching a course called Animals and Philosophy, which critically engages with the place of animals in historical and contemporary philosophy.

12:30 pm - 1:30 pm: Lunch

1:30 pm-2:30 pm

Session 3: Medical Ethics & Tech

Chair: Clarissa Busch

Should an Algorithm Decide if I get Antibiotics? Algorithmic Decision-Makers in the Hospital and the Virtues of Acknowledged Dependence

Charles Freiberg, University of Central Florida

Annette Rid and David Wendler argue for the ethical desirability of developing and implementing an algorithm that clinicians could use to predict incapacitated patients' treatment preferences and assist, supplement, or possibly replace surrogate decision makers (Rid and Wendler 2014a and 2014b; Wendler et al. 2016). Rid and Wendler's moral argument has been maintained by others, who have modified their proposed computational methods (e.g., Earp et. al. 2024). The core argument for the development and use of these algorithms is that they will alleviate surrogate's difficulty in making decisions for incapacitated patients, both the difficulty of determining what the incapacitated patient would want and the difficulty of taking responsibility for this decision. In this paper, I argue that there are virtues required for deciding for another and that the introduction of this type of algorithmic decision-maker will disrupt these virtues. I draw on a strain in the ethics of technology, which analyzes the way technologies mediate practices and as such the way technologies can support or disrupt the virtues (e.g., Borgmann 1984 and 1995; Reijers and Coeckelbergh 2020; Reijers and Gordin 2018; Verbeek 2011). Moreover, I argue that the justifications for developing and implementing these algorithms unintentionally celebrate this disruption. My analysis calls into question whether Rid and Wendler's initial ethical goals are ones we should seek through technological means. It also recontextualizes the debate about these algorithms in terms of the ethics of technology and not simply in terms of bioethics, which has so far dominated the debate.

Charles Freiberg (he/him) is an Assistant Professor of Philosophy at the University of Central Florida. He received his PhD in Philosophy from Saint Louis University, where he wrote a dissertation on educational implications and potential responses to generative AI. He works primarily on issues involving the ethical, social, and political implications of technology and has previously worked on questions related to the implications of AI for universities and the military.

Promises & Perils of Neurotechnology in Clinical Psychiatry

Kathryn Petrozzo, Illinois Institute of Technology

Neuroscience technology (neurotechnology) broadly encompasses the devices and techniques used to understand, assess, and intervene on processes occurring within the brain and nervous system. In clinical psychiatry, neurotechnology has been used primarily as a means of intervening upon psychiatric

conditions as a means of treatment. For example, deep-brain stimulator devices, which are utilized to send electrical impulses through electrodes to affect brain activity, have been successfully used in cases of Parkinson's and Obsessive-Compulsive Disorder. While the clinical applications of neurotechnology are promising means of treatment, there are serious concerns about attempting to use this technology for diagnostic purposes. Primarily, I am concerned with how this technology is being used to not only assert the presence of a mental illness, but also, assert overreaching claims that the presence of a mental illness implies commitments to certain notions of intentionality and responsibility. Notably, that the diagnosis of a mental illness implies that one is less likely to be able to control their behaviors, thus, they are less responsible, and in turn, less likely to be rehabilitated since it is "all in their brain." The danger is when these claims are instantiated in the court of law to justify discriminatory behaviors towards individuals with mental illnesses. Calling upon the literature on underdetermination and scientific expertise, I aim to demonstrate that there is a significant divide between the clinical applications of neurotechnology and the utilization of neurotechnology in legal settings.

Kathryn Petrozzo, Ph.D., is an Assistant Professor of Philosophy in the Department of Humanities at Illinois Institute of Technology and an affiliate faculty member of the Center for the Study of Ethics in the Professions. Prior to her appointment, she was a Visiting Assistant Professor at Oakland University. Kathryn is passionate about working collaboratively to help mitigate discrimination against individuals with mental illness, particularly in carceral settings.

2:30-2:45 pm: Coffee Break

2:45-4:15 pm

Session 4: Decolonizing AI

Chair: Bjørn Ralf Kristensen

Whose Knowledge Counts? Refusal, Embodiment, and the Ethics of Decolonizing AI

Gayatri Raman, Indiana University

Mainstream debates in AI ethics foreground algorithmic bias and governance but often neglect how embodied expertise and legitimacy shape ethical critique. Drawing on an interview study with blind artists in the United States, this paper shows that the stakes of generative AI extend beyond bias to materiality, authorship, and vulnerability. Participants engaged creatively with AI image-generation tools but remained critically attentive to outputs that lacked texture and material presence, questioning whether sighted audiences would interpret imperfections as mistakes or intentional style. These moments highlight how creative agency is mediated or erased when embodiment is discounted. Building on these findings, the paper turns to an ongoing study with disabled and Dalit artists in India, where refusal emerges as an epistemic stance that challenges technological imposition while asserting legitimacy for marginalized knowledge

systems. In these contexts, refusal and embodiment become strategies for contesting the exclusion of disabled and subaltern creative practices in AI-driven cultural economies.

Decolonising AI ethics must move beyond narrow calls for dataset diversity to recognize embodied and marginalized practices as foundational sites of epistemic justice. By theorizing refusal and embodiment, this project reframes the promises and perils of AI and foregrounds whose knowledge and practices are valued in philosophy of technology

Gayatri Raman is a 4th year PhD student at the IU Indianapolis human-centered Computing department working under Dr. Erin Brady. Her work is on accessible creativity and the intersection of embodiment identity and accessible creativity in the Global South. Our current focus is on the creative practices of disabled and marginalized creators from the non-west and how their experience of disability affects their

stance towards AI usage or non usage in their work and the contrast with able non-marginalized creators in the same economies.

Does AI need to be “decolonized”?

Tom Buller, Illinois State University

The following guiding principles are found in many AI regulations and policies.

Ethical and Responsible AI

Privacy, protection, fairness, safety, security, and response for human autonomy and decision-making will be important values in the implementation of the strategy.

Cultural preservation and contextualization

AI systems will be developed that are enriched with Kenyan cultural values and that preserve and promote the nation’s cultural heritage and ensure contextual relevance to local needs and contexts.

The argument has been made that the principles are informed by different epistemologies and values: whereas the first frames AI in terms of its impact on individual rights, the second reflects a more collectivist perspective that privileges culture and community. One reason for claiming that AI is “colonizing” is that until recently the technology has been developed almost entirely in the Global North. Accordingly, the algorithms, data sets, large language models (LLM) have reflected the values and languages of the Global North. A second reason is that the nature of AI privileges an idea of intelligence that emphasizes reason and abstraction, rather than the flexibility and ingenuity of human general-purpose intelligence. Discussions of the harms that AI might cause to society usually focus on harms to the individual. A contrasting perspective is to frame a person in terms of their relationships to others, a perspective that some have argued to be a feature of Ubuntu - an influential moral system in Africa exemplified by the phrase, “I am because we are.”

Tom Buller is Professor of Philosophy at Illinois State University. His main research and teaching interests are in Philosophy of Mind and Neuroethics. His research has appeared in a number of journals and edited collections, including Neuroethics, Bioethics, and the Journal of Medical Ethics.

Maqāṣid al-Sharī‘a as a Framework for Ethical AI: The Perspective of Humanities on Technology and Human Dignity

Soleh Hasan Wahid, State Islamic University of Ponorogo, Indonesia

This study explores how the classical Islamic framework of maqāṣid al-sharī‘a (the higher objectives of law) can be reconstructed to address the ethical, epistemic, and political challenges posed by artificial intelligence (AI). While contemporary AI ethics discourse draws primarily on deontology, consequentialism, virtue ethics, and contractarianism, it often neglects non-Western epistemologies and religious traditions such as Confucianism. Maqāṣid al-sharī‘a offers a complementary lens, grounded in the principles of preservation (ḥifẓ) of life, intellect, religion, lineage, and property, but capable of dynamic expansion in response to historical and technological change.

Drawing on recent AI incidents, from algorithmic discrimination and misinformation to cognitive manipulation and data extractivism, this paper demonstrates how maqāṣid can serve as an evaluative tool that integrates ethical deliberation, systemic risk assessment, and collective welfare. I argue for an expanded conception of ḥifẓ that includes freedom (ḥurriyyah), dignity (‘ird), environment (bī‘a), and information integrity (ma‘lūmāt), thereby situating AI governance within broader civilizational and ecological concerns.

Methodologically, this framework draws from classical jurists (al-Ghazālī, al-Shāṭibī), modern reformers (Ibn ʿĀshūr, al-Fāsī), and contemporary systems theory (Auda), integrating them with risk typologies from MIT and AIAAIC. The resulting approach offers both a comparative ethical dialogue with established theories and a practical rubric for AI design, policies, and education.

By positioning maqāṣid al-sharīʿa as a living ethical paradigm, this paper seeks to bridge Islamic jurisprudential insights with global debates on AI while opening space for plural epistemologies in the governance of emerging technologies.

DTC CK 590 (5th Floor)

9:15-10:45 am

Session 1: Linguistics Art, and LLM

Chair: Elisabeth Hildt

"Beyond Thinking: excess and meaning in thought and its others"

Mark Walter, Aurora University

Descartes famously posits the necessity of having a prior idea of an infinite being in order to recognize our own limitations as thinking things. This can be considered as a certain kind of necessary "excess" to thought itself, and perhaps also as a condition of what we might consider to be specifically human thought, linked as it appears to be to desire and structures of meaning. Centuries later, Emmanuel Levinas similarly invokes an excessive alterity – Autrui, or the other person – as the ground for thought and meaning. This paper examines the phenomenon of AI in light of this concept of thought's excess in order to frame comparisons between human thinking and what seems to appear in Large Language Models, as well as what might be possible in Artificial General Intelligence. At issue will be the question of what "excess" – something fundamentally beyond the grasp of thinking – means to the structure of meaning in human thought, and whether or not the way of being proper to AI systems will permit a workable model of it.

Dr. Mark Walter, Associate Professor of Philosophy at Aurora University. Dr. Walter received his PhD in philosophy from DePaul University, Chicago. He works on aesthetics, the history of philosophy, and contemporary continental thought. Dr. Walter is interested in the topics of artificial intelligence and artificial life, especially with regard to what these can tell us about the nature of human existence.

When should We Believe LLM Outputs? A Testimonial Theory

James Gillard, University of Texas at Austin

Conventional advice about LLM's suggests that their outputs should never be believed without independent corroboration. However, it seems that we can sometimes take LLM testimony at face-value and believe it. When I ask a question like "What is the capital of Lesotho?" most out-of-the-box LLMs will reliably generate the true answer "Maseru." In certain cases, I argue that there is no problem with believing the output. All the same, with more complex questions LLM's are notoriously unreliable: they frequently "hallucinate," generating easily falsifiable responses. How, then, can we explain the fact that sometimes it is justified to believe an LLM and other times not? When exactly is it justified to believe an LLM without corroboration from an outside source?

In this paper, I appeal to the epistemology of testimony to make headway with these questions. I argue that one theory (non-reductionism) has promise for explaining when and why we ought to believe LLM outputs. However, non-reductionism must be modified in a number of ways in order to address the LLM case. Moreover, in order to explain the correct attitude to take when we are not justified in believing LLM

outputs, I will appeal to a broader set of attitudes than epistemologists typically have in mind. I argue that in cases where LLMs are prone to hallucinate, we are justified in “considering” their outputs without being justified in believing them. Keeping such outputs in mind is justified in a way that simply believing them is not.

James Gillard is a PhD candidate in Philosophy at the University of Texas at Austin. His research focuses on moral epistemology, epistemology, and the philosophy of artificial intelligence. His dissertation investigates how agents can form justified beliefs when their sources of knowledge are indirect, opaque, or technologically mediated. In particular, it addresses questions such as whether non-perceptual mental states (e.g. hunches) may justifiably inform our deliberation and when we are justified in trusting the outputs of artificial intelligence. In addition to this contemporary work, he maintains a secondary specialization in ancient philosophy, especially Plato and the moral psychology of Socratic inquiry.

AI Chatbots and the Paradox of Fiction: (How philosophy of fiction can shed light on AI-human relations)

Sydney Harvey, Illinois Institute of Technology

Abstract: This paper introduces "The chatbot companion paradox," which draws a parallel to the paradox of fiction. The chatbot paradox highlights the inconsistency of people feeling companionship towards AI chatbots while simultaneously knowing that these machines lack emotionally complex minds. It explores philosophical concerns about the mental capacities of large language models (LLMs) and concludes that current LLMs are unlikely to reciprocate as genuine companions due to their reliance on word prediction rather than true understanding or emotional intelligence. Ultimately, the paper proposes a solution to the chatbot paradox by arguing that companionship does not entail a belief in the emotional complexity of the object of affection. Instead, it suggests that people project empathetic emotions onto objects, and the chatbot acts as a mirror for these projections, providing an external layer of self-affirmation.

Sydney Harvey is currently completing my dissertation in the areas of artificial intelligence, mind, and aesthetics under the supervision of Jesse Prinz and Noel Carroll at CUNY Graduate Center. Her research interests consider the correlations between environment, embodied emotion, perception, and ethical engagement in both human and machine minds. She is currently a pre-doctoral research fellow with the Fay Horton Sawyer Center for Ethics in the Professions at the Illinois Institute of Technology (IIT). The Sawyer fellowship consists of a year-long research and teaching appointment. Sydney has been published in *Hyperallergic* with a piece called, "Visiting MoMA While Black," as well as an academic paper in *Film and Philosophy Journal* titled, "A Reflection of the Sun."

10:45 am - 11:00 am: Coffee Break

11:00-12:30 am

Session 2: Law & Jurisprudence

Chair: Kattrhryn Petrozzo

Against AI Jurisprudence: Large Language Models and the False Promises of Empirical Judging

Dasha Pruss, University of Illinois, Chicago

As hype around the transformative effects of large language models (LLMs) has taken center stage in popular culture, some judges and legal scholars have suggested that LLMs have the potential to improve the objectivity of judicial decision-making. Proponents argue that using LLMs to find empirical ‘evidence’ of legal text’s meaning can reduce the role of judges’ subjective choices, ensuring that judicial rulings faithfully reflect the people’s understanding of legal rules, and grounding legal interpretation in a sophisticated empirical investigation of real language use in social context. To the contrary, we argue that

LLM jurisprudence underscores the discretionary decisions required to infer ordinary meaning; highlights the inescapable reality that the meaning and application of legal terms is inherently normative; and demonstrates the lack of democratic legitimacy of crowdsourcing legal meaning. We argue that the feature of LLMs that makes them so seductive for legal interpretation – their potential ability to approximate ‘ordinary’ people’s understanding of legal text – reveals the political illegitimacy of empirical judging. We conclude with recommendations and warnings for practitioners in this space.

Dasha Pruss (she/her) is an Assistant Professor of Philosophy at the University of Illinois Chicago and a Faculty Associate at the Berkman Klein Center for Internet & Society at Harvard University. She received her PhD in History & Philosophy of Science from the University of Pittsburgh in 2023. She is currently on leave from George Mason University, where she was an Assistant Professor of Philosophy and Computer Science from 2024-2025. Before that, she was a postdoctoral fellow at Harvard University, where she was jointly a fellow at the Berkman Klein Center and a postdoctoral fellow in the Embedded EthiCS program. Dasha’s research critically examines the societal implications of AI/ML systems, with a focus on algorithmic decision-making systems used in the US criminal legal system. She has written about recidivism risk assessment instruments, predictive policing tools, electronic monitoring, and other carceral technologies. In 2024, she organized Prediction and Punishment: Cross-Disciplinary Workshop on Carceral AI, which brought together scholars and activists from around the world to address technologies designed to police, incarcerate and surveil human beings.

The Application of the Rule of Law: Technological Assistance and the Ontic/Ethic/Epistemic Paradigm

Dennis Kamalick, Chicago Police Force Training Academy

Despite occasional flaws, there remains a reasonable amount of trust in data systems. Yet there persists a "privacy paradox": whereby the convenience of immediate access masks the danger of user vulnerability, while supporting an illusion of user security. There are many dynamic elements of society. It is in the nature of those in certain professions to stand for (i.e. to promote, support and defend) their respective societal element. Law enforcement agencies function in this maintenance role by preventing and investigating crime, by apprehending criminals, by responding to emergencies. Toward that end, such agencies rely on technological assistance for communication, data analysis, dispatch and documentation, surveillance, global positioning, infrared three-dimensional imaging, advanced databases and other tools. Considering law enforcement technology it is well to remember that technology itself is a useful but a neutral tool. Imposing an algorithmic bias, or adjusting a search to satisfy a prejudiced mindset, is caused, not by the sophisticated technological procedure for being so complex, not by "Big Tech" for inventing the sophisticated and complex procedure, but it is caused by its abuser. In approaching the task of applying the law, the law enforcement official should nurture a profound respect for the rule of law in its fundamental existential reality as the foundation of civil society, and expressed in the local social community. The officer then upholds that rule as an active participant in that local social community. In so doing, the law enforcement officer provides the community with a service; a service inspired by a positive disposition of respect for the fellow members of the community whom the officer serves and among whom the officer lives. Thus, the law enforcement officer honestly acknowledges as true that one indisputable fact that produces evidence to inspire impartially honest reasoning toward a rational conclusion, uncompromised by assumption, conjecture or bias.

A Chicago native, Dennis spent thirty years in a religious order, during which he was able to achieve a terminal degree in Philosophy at the Pontifical Gregorian University in Rome. While an adjunct lecturer with City Colleges of Chicago, Dennis taught humanities disciplines to officers and civilian staff in evening classes at the Chicago Police Training Academy. He joined the Police Department in 2001, spending sixteen years as an investigator of allegations of police misconduct. After serving two years in a clerical position at Public Safety Headquarters, Dennis has been, since 2021 an instructor with the

Chicago Police Department's Training and Support Group, teaching classes in "Theories of Crime and Crime Indexing", "Victims' Rights", "Neurobiological Factors Contributing to Trauma", "Police and Community Relations", and others. Currently he is developing a series of what might be stylized as "vignette courses", or "brain teaser challenges", or "invisible flash card shout-outs", all centered on "Critical Thinking", to be eventually incorporated into the Academy curriculum.

Why Copyright Must Remain Human-Centric: An Ethical Analysis of AI and Copyright

Eylül Erva Akin, Bocconi University, Milan, Italy

The distinction between creativity and copyrightability is profoundly an ethical question. Creativity can emerge from humans or machines. Copyright, however, is a legal term built on recognition, responsibility and rights. To extend copyright to AI-generated creations would mean treating machines not as tools, but as potential authors, causing ethical, economic and societal risks. The question is thus not only “can AI create?” but “should AI be recognised as a rights-bearing author?” AI ethics already faces similar dilemmas whether machines can have moral status, whether their autonomy entitles them to dignity and whether anthropomorphism risks attributing humanlike agency where none exists. If authorship requires autonomy, intentionality and responsibility, then copyright protection for AI creations risks severing the link between rights and accountability. This paper argues for a human-centric approach: copyright protection is justified only where the human involvement is decisive. This position rests on three normative grounds. First, from a rights-based perspective, responsibility gaps show why copyright must remain tied to human dignity and authorship, ensuring accountability and recognition of labour. Second, in utilitarian terms, distributional risks demonstrate that granting copyright to AI-generated creations would concentrate economic benefits in corporations while diminishing incentives for human creators. Third, through virtue ethics, cultural flattening reveals how recognising AI as an author would erode creativity as human self-expression and meaning-making, reducing artistic activity to automated, commodified outputs. No matter which ethical framework adopted, this paper concludes that copyright should remain human-centric.

Eylül Erva Akin recently submitted her PhD in Law at the University of Milan and is awaiting her viva. Her work focuses on AI regulation, copyright law and legal and ethical implications of technology. She is a Research Assistant at Bocconi University within the Digital Constitutionalism (DIGCON) Project, examining how fundamental rights shape AI regulatory frameworks. She teaches AI and digital transformation law at Bocconi University and the University of Milan. She is a qualified lawyer (Istanbul Bar) and holds an LL.M. in IT and IP Law (Leibniz University Hannover & Queen Mary University of London).

12:30 pm - 1:30 pm: Lunch

1:30 pm-2:30 pm

Session 3: Responsibility

Chair: Kelly Laas

Beyond Good and Evil: Thinking AI within the Human Condition

Jurgita Imbrasaite, University of Bonn

The ethical debate around artificial intelligence has become both urgent and ubiquitous, yet it is still framed predominantly in terms of a binary logic of “good” and “evil.” Concerns about the trustworthiness of AI—foregrounded by the EU AI Act and debated across contexts from care robots to autonomous vehicles and large language models—are often articulated through anxieties about harm, deception, or

loss of control. Such framing, however, risks reducing a profoundly philosophical problem to a superficial moral schema, echoing popular fear rather than critical reflection.

Modern philosophy has long destabilized the moral categories that still shape AI ethics today. Since Kant, and more radically through Nietzsche and post-structuralist thought, the question of ethics has been revealed as emptied of stable foundations, while psychoanalytic readings (e.g. Jacques Lacan, Alenka Zupančič) underscore its paradoxical structure. In this light, it becomes insufficient to treat AI merely as a potential bearer of good or evil outcomes. Instead, we must consider AI within the entangled condition of human life and technology, where responsibility, agency, and meaning are co-produced. Hannah Arendt already observed that modern technology transforms the relation between human activity, world-building, and technical systems in ways that exceed instrumental reasoning or moral clarity. Similarly, Gilbert Simondon emphasized the interwoven genesis of technological objects, inseparable from human development and maintenance.

This paper proposes to rethink AI ethics as negotiation within a shared field of human–AI co-action, moving beyond moral binaries toward a philosophy adequate to our entangled, technologically mediated condition.

Dr. Jurgita Imbrasaite (she/her) is a Senior Research Fellow at the Institute of Philosophy at the University of Bonn. She is conducting her research on “The *Conditio Humana* in the Age of AI” as part of the ‘Desirable AI’ project and is the founder of the AI-ALLY initiative. Her philosophical work investigates the discursive and epistemic conditions of modernity, with particular attention to how the human condition evolves under new regimes of knowledge, power and technology. She received her interdisciplinary doctorate in 2015 from Ruhr-University Bochum, where she examined the shift from sovereign societies to the knowledge-driven modern age through the theories of Jacques Lacan and Michel Foucault. In recent years she has expanded her scope to include political and philosophical questions about the technological condition of modernity. Her current research focuses on Hannah Arendt's political philosophy, the modern labor paradigm, and the epistemic conditions for human-technological co-activity. Before joining the University of Bonn she was a research associate at the Institut für Geschichte und Zukunft der Arbeit (IGZA) and a lecturer at the University of Europe in Hamburg.

The Limits of AI Understanding and What they Mean for Users

Michael Zahorec, Florida State University

I clarify what understanding AI is with two foundational distinctions: interpretable AI versus explanations of AI and internal-process explanations versus black-box explanations. These clarify the ways we can know about how AI systems operate, and are both practically and normatively significant. I demonstrate their normative significance by comparison to the moral significance of motives. As an action's moral status is often partially determined by properties that cannot be observed behaviorally, so too does what occurs inside the black box matter. I demonstrate this with a simple example involving a hiring algorithm with multiple internal decision pathways, one of which is discriminatory. I show systematic limitations of black-box explanation—limitations that cannot be overcome with better experimental design or larger datasets. I then use these same foundational distinctions to frame the most pressing current challenges for understanding contemporary AI, focusing on generative AI and on LLMs in particular. Interpretability is presently infeasible, self-explanations and other black-box explanations are unreliable, and internal-process explanation efforts such as mechanistic interpretability currently face serious challenges. I argue we should be cautiously optimistic about overcoming these challenges but that, in the meantime, certain forms of trust in and reliance on complex AI systems are ruled out—especially in normatively significant contexts like hiring, legal proceedings, and medical decision-making. I conclude by spelling out a recommendation for responsible AI use: that we should generally treat AI systems as idea

generators, as opposed to other paradigms like AI-as-expert, AI-as-decisionmaker, or AI-as-neutral-information-processor.

Michael (Mike) Zahorec is a PhD candidate in Philosophy at Florida State University, where he will defend his dissertation in March 2026. His dissertation develops a permissive pluralist realism about scientific kinds, contributing to the contemporary literature in philosophy of science while drawing on ideas from Kant and Wittgenstein. While at FSU, Mike completed an MS in Computer Science, with a thesis that developed methods for evaluating the quality of large language model (LLM) explanations. Mike's current work concerns our efforts to evaluate, to explain, and to responsibly use AI. He is especially interested in how these bear on each other: that is, the connections between the epistemology and the ethics of AI. His interdisciplinary approach combines philosophical analysis with technical AI research and development, including industry research on LLM evaluation with Humana's Responsible AI team. Mike enjoys cooking, thrift shopping, and going on adventures with Elly (his spouse) and Bingo (their dog). He also enjoys reading – especially biographies and anything written by Anscombe or Wittgenstein.

2:30-2:45 pm: Coffee Break

2:45-4:15 pm

Session 4: Agency

Chair: Kelly Laas

Agency Collapse in Engineering

Nick Treanor, University of Edinburgh

This talk will focus on the phenomenon of agency collapse in engineering – identifying what it is, giving examples of it, arguing that it involves a specific kind of failing that is both ethical and epistemic at once, and showing how various technologies in widespread use in engineering tend to promote it. The talk will centre on recent, real-world examples of disasters (large and small), primarily the case of the Grenfell Tower Fire in London, England in 2017, and cases in civil, structural and fire engineering drawn from recent CROSS reports. (CROSS is Collaborative Reporting for Safer Structures, an international confidential reporting system wherein professionals working in the built environment report on fire and structural safety issues they've encountered to share lessons and improve safety). In each case, I will outline the particular example and show why it should be understood as an instance of agency collapse engendered by technology. The goal will be not only to deepen our understanding of the ethical, epistemic and practical role of technology in engineering failure, but to offer a diagnosis of a widespread problem that has the potential to suggest fruitful interventions and more apt design and use of the technologies in question. Although the talk will focus on civil, structural, fire and construction engineering, the phenomenon of agency collapse and the technologies involved are pervasive across many industries. The analysis and proposals offered, therefore, are intended to be of general interest.

Nick Treanor (he/him) is a Reader in Philosophy at the University of Edinburgh and a research affiliate with the Edinburgh Futures Institute. Previously, he was the Newton Trust Lecturer in Philosophy at the University of Cambridge and a Fellow of Churchill College. He's originally from North America, where he did an undergraduate degree at Queen's University (Canada) and a PhD at Brown University (USA). His work focuses on theoretical issues concerning how to understand quantities of knowledge, understanding and ignorance, as well as on the ethical and epistemic dimensions of engineering. He works closely with senior industry leaders in civil, structural, fire-safety, forensic and construction engineering. The aim of this collaborative research is both theoretical – to better understand important aspects of thought and action – and practical – to improve engineering, how it is conceived, taught,

governed, and practiced. As part of this work, Nick is co-supervising, with engineers, a PhD student working on the epistemology of construction product certification, as well as a Master's student working on the normative identity of engineering.

Why Machines cannot be Full Moral Agents

Dane Leigh Gogoshin, University of Wisconsin

Why Machines cannot be Full Moral Agents The ongoing debate about artificial moral agency (AMA) tends to focus on the question of moral responsibility. If machines cannot be (legitimately held) morally responsible for their actions, the thought goes, then they are not moral agents. In this paper, I argue against this line of thought. Due to the enormous diversity and flexibility of moral responsibility theorizing, especially its social practice-centered (Strawsonian) orientation, it is too easy to argue one way or the other. According to my own practice-based view of responsibility, responsibility is not the apex of moral agency; moral autonomy is. Moral autonomy is thus the right benchmark for full AMA. It involves a cluster of capacities that, while immune to the many controversies that plague responsibility theories which preclude AMA, are not available to conceivable those machines.³ Among these capacities are an ability to meaningfully grasp moral considerations (as per Kauppinen 2024) and to respond directly to them, and the ability to negotiate (possibly construct) the normative landscape (as per McGeer 2019). While this conclusion – that machines cannot be full moral agents – does not entail either that machines cannot be responsibly deployed or attain some degree of moral agency, it does entail that they should not replace human beings in contexts where moral decision-making is required.

Dane Leigh Gogoshin (PhD Helsinki) is a 2024-2026 visiting assistant professor of philosophy at the University of Wisconsin-Madison's Information School. She is currently teaching data ethics & policy courses and a course on moral responsibility & AI to STEM majors. Her primary research interests are on autonomy, moral responsibility, moral autonomy, and related social and legal practices, and their relationship with data technologies (especially AI and social robotics).

