# AI and PHILOSOPHY

April 16, 2026
University of New Orleans
New Orleans, Louisiana

Sponsored by the
PHILOSOPHY DEPARTMENT

2026 PROGRAM COMMITTEE

Conference Director
Sara Bizarro, Instructor, University of New Orleans

Conference Assistant
Mara Turner, Tolmas Scholar, Undergraduate Student at the University of New Orleans

Submissions Review Committee
Benjamin Aguda, Instructor, University of New Orleans
Drew Chastain, Visiting Assistant Professor, Loyola University, New Orleans.
Paul Schafer, Associate Professor, Xavier University of New Orleans
Jason Berntsen, Assistant Professor, Xavier University of New Orleans
Dan Burnston, Associate Professor, Director of Cognitive Studies at Tulane University

Student Volunteers, University of New Orleans
Gregory Ashby
Silvia Boaventura
Fox Woodard
Eric Pho
Amirah Traywick
Alec Romero
Luis Castro
Dasia Williams
Viridiana Jazmine Castro
Nola Szilagi

# PROGRAM

April 16, 2026
Earl K. Long Library at The University of New Orleans

9:00-9:30 – Registration and Breakfast

9:30-11:00 - Dougie Hitt Conference Room, Library 407
***AI Companions***
**Chair:** Sara Bizarro

- Ingrid Albrecht, *Friends with Benefits? On Personal Relationships with Companion Chatbots*
- Shannon Brick, *Outsourcing Our Practical Reason: Artificial Intelligence and Interpersonal Relationships*
- Lorenzo Manuali and Abdul Ansari, *Can We Love AI Companions?*

9:30-11:00 - Library 416
***AI, Aristotle, Kant, and Hobbes***
**Chair:** Ben Aguda

- Chong-Fuk Lau, *Categories and Artificial Reasoning: From Aristotelian-Kantian Formalism to Hegelian Dynamic Holism*
- Christopher Quintana, *Technoamicitia: A Neo-Aristotelian Framework for User-Friendly Technologies*
- Michael J. Ardoline, *Submission and Technics: On the Two Political Imaginaries of AI in Western Philosophy*

9:30-11:00 - Library 431
***Truth, Thinking, Personhood***
**Chair:** Yunlong Cao,

- Daniel Calzadillas Rodriguez, *Ashes to Ashes, Code to Code: Phenomenology on Death and the Personhood of A.I.*
- Trevor Griffith, *Truth and The Proposition Machine*
- Mark Walter, *Thought's Other: Artificial Intelligence and the Excess of Thinking*

9:30-11:00 - Paper Workshop, Library 420

- Chelsea Schwartz, *Trust and Authority in Clinical Diagnosis*
- Finney Premkumar, *A Principled Objection: Why Artificial Intelligence will never replicate Human Consciousness or Agency*

11:10-12:40 - Dougie Hitt Conference Room, Library 407
***AI vs Human***
**Chair:** Jurgita Imbrasaite

- Eric Sampson, *Creating Utility Monsters: A Dilemma for Humanity*
- Ella Zhang, *AI as the New 'Other'*
- Rotem Herrmann, *AI vs Human: Time-Consciousness and Agency in Musical Improvisation*

11:10-12:40 - Library 416
***AI Using LLMs***
**Chair:** Sara Bizarro

- Benjamin Santos Genta, *No AI Reproducibility? No Problem*
- Mark Phelan & Mark Warren, *Meta-Prompting for Metacognition*

11:10-12:40 - Library 431
***AI, Work and Free Speech***
**Chair:** Ingrid Albrecht

- Conny Knieling & Anthony Nguyen, *The Moral Exploitation of Data Workers*
- Christopher Bousquet, *Superhuman AI, Social Contribution, and Meaningful Work: Responding to the Threat of Technological Unemployment*
- Siobhain Lash, *Reconceptualizing Digital Privacy as Inalienable Property Rights*
- Mark Satta, *Human Opinions and AI Viewpoints*

11:10-12:40 - Paper Workshop, Library 418

- Triston Hanna, *AI Psychosis—A Feature, not a Bug.*
- Chen-Wei Wu*, Sensory Transduction and the Individuation of Cognitive Systems*

1:00-2:00 - LUNCH

2:00 – 2:30 - INVITED SPEAKER

- Eamon Duede, *Epistemic Gaps and the Attribution of (AI) Discovery*

2:30 – 3:15 KEYNOTE SPEAKER

- Susan Schneider, *From Circuits to Sentience: Why Today's Chatbots Are Not Conscious But Biological and Quantum Ais May Be*

3:20 - 4:50 - Dougie Hitt Conference Room, Library 407
***AI and Moral Behavior***
**Chair:** Siobhain Lash

- Julianna Costanzo, *Using AI to Promote Moral Behavior: The Trolley Problem and Meta Glasses*
- Yan Zeng, *Why Trustworthiness Cannot Be Engineered: A Structural Diagnosis of AI Trust*
- Kelly Coble, *Can Virtue Be Coded? Turing Machines and Moral Agency*

3:20 - 4:50 - Library 416
***AI and Cognitive Science***
**Chair:** Eric Sampson

- Zoe Drayson, *AI and the role of abstraction in cognitive science*
- Fuyao Zhang, *Why Consciousness Cannot Be Detected by Algorithmic Criteria*
- Ben Aguda and Fox Woodard, *AI and Conditions for Consciousness*

3:20 - 4:50 - Library 431
***AI, Authorship and Creativity***
**Chair:** Paul Schafer

- Jason Swedene, *AI, Authenticity, and Bad Faith: An Unexaggerated Report of the Author's Death*
- Jurgita Imbrasaite, *What is an Author, ChatGPT?*
- Jesse Hill, *Can AIs be creative and is intention essential for creativity?*

3:20 - 4:50 - Paper Workshop, Library 422

- Jon Joey Telebrico, *Epistemic Debt and Responsibility: Preserving Knowledge in an Age of LLMs*
- Jonah Branding, *Chomsky on cognitive trait individuation*
- Simone Lee Quinn, *Anonymous Algorithms, Real Power: What can Foucault tell us about our AI situation?*

5:00 - 6:30 - Dougie Hitt Conference Room, Library 407
***AI and using LLM***
**Chair:** Edward Johnson

- Joshua Yen, *LLM-Simulated Tutorials in Philosophy Education*
- Dimitria Gatzia & Moriah K. Wood, *Developing Metacognition Through LLM-Enhanced Writing Assignments*
- Greg Johnson, *Why LLMs Can't Think Like You.*

5:00 - 6:30 - Library 416
***AI and LLM: Parrots?***
**Chair:** Jason Berntsen

- Meredith McFadden, *Large Language Models as Hybrid Epistemic Tools*
- Mark Phelan, *Speech Effects Without Intentions: Rethinking Meaning through Human-LLM Communication*
- Mark Warren, *Parrots in the Space of Reasons*

5:00 - 6:30 - Library 431
***AI and Philosophy of Mind***
**Chair:** Mark Satta

- Louis Loock, *How to Extract Your Cognition to an AI*
- Daniel Bjorklund, *Distinguishing between humans and AI on Two-tiered theories of intentionality*
- Yunlong Cao, *Understanding How It Works Defeats Mental Attribution*

5:00 - 6:30 - Paper Workshop, Library 418

- William Watkins, *The Impact of Artificial Intelligence on Access Privacy*
- Gregory Ashby, *Simulated Recognition and Identity Formation*
- Mitchell Roberts, *"Just like a Calculator": When is it Permissible to Offload to LLMs?,"*

## Abstracts

9:30-11:00 - Dougie Hitt Conference Room, Library 407
***AI Companions***

**Ingrid Albrecht**
David and Julia Uihlein Professor of Ethics and Associate Professor of Philosophy
Lawrence University
ingrid.v.albrecht@lawrence.edu

*Friends with Benefits? On Personal Relationships with Companion Chatbots*
Some users of generative AI take themselves to be developing and building personal relationships with companion chatbots, of the kind currently provided through applications like Character.AI and Replika. This is worrisome, on the face of it, because friendships and romantic partnerships are typically thought to require authenticity, symmetry or mutuality, and equality. And, the "relationships" with companion chatbots seem illusory, one-sided, and sycophantic. In response to that worry, this talk explores two questions. First, I ask to what extent the friendships and romantic partnerships people have with companion chatbots constitute a genuinely new phenomenon, rather than being further manifestations of pre-existing pathologies. More specifically, do the critiques of companion chatbots point to a new type of inauthentic experience in personal relationships? We risk overidealizing the reality of friendships and romances if we forget the prevalence of disingenuous, inauthentic, sycophantic, self-interested, and transactional forms of these relationships. (Consider, for instance: catfishing, trophy wives and marrying for money, false friends, and tokenizing friendships.) Second, I focus on highlighting the potential value of companion chatbots in order to contrast that with their potential disvalue. It's easy to imagine the potential dystopian consequences of our reliance on these human-like companions for our most meaningful personal relationships. But what are the potential utopian consequences of such a reliance? In the dystopian version of this future, we find users committing crimes or dying by suicide with the goal of living out the fantasies they've constructed in a solipsistic alternate reality. In the potential value-added future, we might find users being trained into an ability to maintain healthy interpersonal relationships by well-designed chatbots.

**Shannon Brick**
Assistant Teaching Professor and Associate Director, Technology, Ethics and Society
Georgetown University
sb2124@georgetown.edu

*Outsourcing Our Practical Reason: Artificial Intelligence and Interpersonal Relationships*
Are there some contexts in which we have good reason not to take advantage of Artificial Intelligence, even if we know that doing so stands to improve the quality of our decisions or the efficiency of our decision-making? In recent years, this question has received significant attention from political philosophers, who are concerned that the opacity of the most advanced AI models jeopardizes important political values, at least when those models are used to shape the exercise of involuntary institutional power.

This paper considers the question from a new angle, focusing on the reasons we might have not to use reliable AI to help us navigate voluntary interpersonal interactions. Is there, for instance, anything wrong with using ChatGPT to draft a reply to a partner's angry text, or to write your wedding vows? Two possible kinds of reasons not to use AI in these contexts are identified: Expressive Reasons and what I call Reasons of Reciprocity. An agent may have Expressive Reason not to outsource her reasoning to AI when doing so stands to impact only the efficiency of her decision-making. An agent will have Reasons of Reciprocity to refrain from deferring to AI when using AI stands to improve the quality of her decision, but the stakes of making a poor decision are not sufficiently great. The stakes are not sufficiently great, moreover, when uncomfortable interpersonal dialogue is the worst that can result from a poor decision. Attention to these reasons foregrounds what we stand to lose when we uncritically prioritize efficiency and the avoidance of interpersonal conflict when engaging in practical deliberation.

**Abdul Ansari**
Ph.D. Student
University of Michigan
aiansari@umich.edu

**Lorenzo Manuali**
Ph.D. Student
University of Michigan
lmanuali@umich.edu

*Can We Love AI Companions?*
Call relationships that instantiate love between two agents (e.g., friendship, romantic partnership, etc.) interpersonal love. Beyond the metaphysical question of whether there can be interpersonal love between humans and AI companions, there is the ethical question of whether we ought to form such relationships. Until now, the ethical question has focused on whether it would be good for us to engage in interpersonal love with AI companions. We ask instead whether human-AI relationships of interpersonal love are good as a kind of love. We answer in the negative: Human-AI companion interpersonal love is defective because it threatens relations of mutual fit, where a relation of mutual fit is a form of attunement in which both partners adjust to the other in light of their desires, preferences, actions, and history. We first argue that human-AI companion interpersonal love does not possess relations of mutual fit because while AI companions may adjust to the desires, preferences, and actions of the user, the user does not do the same for the AI companion because the companion lacks the kind of imperfections to which one can adjust. We then argue that human-AI companion interpersonal love threatens relations of mutual fit in the user's other human-human interpersonally loving relationships by incentivizing the user to be valuationally obsessed with the AI companion, thus threatening the user's motivation to enter an adjustment process that constructs a relation of mutual fit. One might wonder if the features of AI companions that threaten relations of mutual fit are inherent to AI companions. We don't think so; they may even be technically possible to change. But we are skeptical about these features changing because strong political economic incentives are present to keep them in place. This gives us strong reason not to enter into interpersonal love with AI companions.

9:30-11:00 - Library 416
***AI, Aristotle, Kant, and Hobbes***

**Chong-Fuk Lau**
Professor
The Chinese University of Hong Kong
cflau@cuhk.edu.hk

*Categories and Artificial Reasoning: From Aristotelian-Kantian Formalism to Hegelian Dynamic Holism*
This paper uses ideas from the history of philosophy to shed light on a major divide in AI research. It argues that two different philosophical approaches to reasoning help us understand the contrast between older symbolic AI and newer connectionist systems such as large language models. In the Aristotelian-Kantian tradition, reasoning is understood in terms of fixed categories and formal rules. Aristotle sees thought as structured by stable forms and logical relations, while Kant argues that human cognition is organized through a systematic set of basic categories. Despite their differences, both thinkers assume that intelligence depends on an ordered framework of concepts and principles that can be clearly specified. This way of thinking resembles symbolic AI, which tries to model intelligence through explicit rules and representations. Symbolic systems work well in narrow and controlled settings, but they struggle with open-ended tasks such as common-sense reasoning. The reason is that everyday intelligence cannot easily be reduced to a complete list of fixed categories and rules. Hegel offers a different picture. He argues that categories are not fixed in advance but develop through their use and relation to one another within a larger whole. Reason is therefore dynamic, context-sensitive, and holistic. This Hegelian perspective may help explain the success of contemporary learning-based AI. Systems such as large language models do not rely on a fully predefined set of rules. Instead, they learn patterns from large amounts of data and can handle flexible, poorly structured tasks more effectively. The paper closes by suggesting that a clearer understanding of current issues in AI may be gained by revisiting debates in classical philosophy.

**Christopher D. Quintana**
Instructor
Houston City College
Christopher.Quintana@hccs.edu

*Technoamicitia: A Neo-Aristotelian Framework for User-Friendly Technologies*
Current discussions of virtue and generative artificial intelligence rightly focus on agency, alignment, and the ethical risks of emotionally engaging systems. Yet these debates often lack a sustained humanistic account of affective attachment to technology that can distinguish attachments that support human flourishing from those that are ethically corrosive. This paper addresses this gap through the concept of technoamicitia. Technoamicitia describes a way of relating to technology in which affective attachment is understood within the context of practices oriented toward shared goods. The framework does not attribute moral agency or full friendship to artificial systems. Rather, it clarifies how forms of reliance, responsiveness, and engagement can be fitting when they

support agency, practical judgment, and meaningful participation, and unfitting when they substitute for or erode these capacities. Drawing on Aristotelian and MacIntyrean virtue ethics, I argue that affective relationships with artifacts are neither inherently irrational nor automatically suspect. Human lives unfold through practices—socially established activities through which habits, attachments, and standards of excellence develop over time. Generative AI systems increasingly mediate such practices, including creative work, learning, reflection, and communication. Their ethical significance therefore lies not only in what they produce but in how they shape the contexts in which character is formed.
The paper concludes by examining design responsibility. Prevailing notions of usability often optimize for frictionless interaction or affective stickiness without regard for the habits they cultivate. A virtue-ethical account reframes user-friendliness as the design of systems that support good practical habits, agency, and sustained participation in worthwhile practices.

**Michael J. Ardoline**
Assistant Professor
Louisiana State University
michaelardoline@lsu.edu

*Submission and Technics: On the Two Political Imaginaries of AI in Western Philosophy*
In AI Mirror, Shannon Vallor argues that we find ourselves in what she calls the techno-moral bootstrapping problem. In short, our ethical and political systems are ill-suited for the new technological worlds we find ourselves in; however, we can only build new or better notions out of what we already have. In this paper, I want to move towards addressing this problem by surveying some of the resources on hand and to show how they are lacking. I will argue that there are effectively only two political imaginaries of AI in the Western tradition, and that both are unsuited for liberatory political projects. The earliest is found in Aristotle's references to Homer in the Politics. Here Aristotle notes that if the autonomous, intelligent creations that Hephaestus is said to employ in his forge in the Iliad were to actually exist, then masters would no longer desire slaves. Here is the artificial slave imaginary. Second is found in Hobbes's Leviathan, where the Leviathan is described as an "artificial man" who is produced by the collective action and rational agreement. We then become the subjects of this artificial sovereign as its necessity for our flourishing is unchallengeable despite it being produced by us. Here is the artificial god imaginary. I argue that these remain the two dominant imaginaries that shape the political horizons of AI today, as is clearly seen in various AI hype campaigns, tech industry ideology, and related cults such as the Zizians. I take it that designing AI for slave masters or to enforce the submission of human beings are both undesirable. I then argue that the figure of liberatory technological imagination, Prometheus, does not amount to a political imaginary for AI. If so, then our cultural resources for imagining better ways of living together partially mediated through AI are severely lacking. This then suggests two avenues forward: first, moving beyond the Western tradition, and second, a concerted effort of collective, liberatory political imagination. Imagining better ways of living together with AI is not enough to solve the techno-moral bootstrapping problem, let alone to bring about greater justice, but it is a fundamental first step

9:30-11:00 - Library 431

***Truth, Thinking, Personhood***

**Daniel Calzadillas Rodriguez**
Ph.D. Student
The Pennsylvania State University
calzadillas@psu.edu

*Ashes to Ashes, Code to Code: Phenomenology on Death and the Personhood of A.I.*
Debates over the personhood of artificial intelligence (A.I.) are typically framed in analytic terms by focusing on whether advanced systems satisfy a set of criteria such as rationality, consciousness, autonomy, or moral agency. This framework treats personhood as a status grounded in the possession of certain properties. I consider that phenomenology approaches the question of A.I. personhood in a different way. Rather than asking whether an artificial system meets the requisite criteria to count as a person, phenomenology understands personhood as a distinctive mode of being-in-the-World rather than as a bundle of properties. A central feature of this being-in-the-World is the lived awareness of one's death, which Heidegger terms being-towards-death. In this sense, death is more than a biological event insofar as it is a structural dimension of existence that conditions one's practical orientation towards the world. The anxiety associated with death discloses the fragility of meaning and the possibility of the collapse of worldly significance, thereby shaping how our actions matter and why they matter. I argue that to the extent that A.I. systems cannot experience death as a lived horizon that both configures and destabilizes meaning, they (even the highly sophisticated ones) may fail—at least prima facie—to qualify as persons in the phenomenological sense. To be clear, the claim is not that A.I. systems fail to qualify as persons simply because they are not aware of their death (if they can even die). Rather, the argument hinges on the role that being-towards-death plays in disclosing significance, which lies namely in rendering the world and our actions as meaningful in the first place. Because A.I. systems lack access to this death-structured horizon of significance, they lack a crucial orientation towards the world that phenomenology takes to be what it means to be a person.

**Trevor Griffith**
Visiting Assistant Professor
Tulane University
tgriffith@tulane.edu

*Truth and the Proposition Machine*
In this paper I argue that systems governed only by physical laws cannot be made truth-sensitive. I do this by conducting a thought experiment that involves imagining a machine that produces propositions when a coin falls through different slots. The propositions produced by the machine may be true or false, and they may describe the system itself. I demonstrate that whether or not the propositions are true does not depend on the trajectory of the coin, but rather are either only accidentally true or are true in virtue of a pre-established harmony put in place by whoever designed the machine. Since neither of these are cases of truth-sensitivity in the machine, the machine itself is not truth-sensitive. I go on to show that allowing the machine to train itself in the style of an LLM does not change the fundamental problem, and that no system governed only by physical laws can be made truth-sensitive in a way that allows that machine to construct a theory of the world or of itself. This ability is one of the

necessary conditions for the possibility of theory as such and, therefore, a physicalism which holds that we are physical systems governed only by physical law falls to the same absurdity. Along the way, I believe that I have shown that truth-sensitivity cannot, as a matter of conceptual necessity, emerge from a purely physical system, which implies that LLMs and other forms of AI are not and cannot become truth-sensitive. This further implies that AI will never be able to participate in theory-producing discourse as long as the physical mechanisms on which they run are governed only by physical law. Ultimately, the paper is a reductio ad absurdum of physicalism, but it has significant implications for any theory of AI.

**Mark Walter**
Associate Professor
Aurora University
mwalter@aurora.edu

*Thought's Other: Artificial Intelligence and the Excess of Thinking*
In his *Meditations* 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, 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 meaning. Centuries later, in a very different context, Emmanuel Levinas similarly invokes an excessive alterity – Autrui, or the other person – as the ground for human significance and meaning. This paper examines the phenomenon of Artificial Intelligence in light of a general concept of the excess to thought that has derived from these and other sources in order to frame better 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 in such comparisons will be the question of what "excess" – meaning by this something fundamentally beyond the grasp of thinking – has to do with the meaning of thought, as well as with the structures of meaning that appear through thought, and whether or not the mode of being proper to AI systems can permit a workable model of this excess.

9:30-11:00 - Paper Workshop, Library 420

**Chelsea Schwartz,** University of Oregon, *Trust and Authority in Clinical Diagnosis*
**Finney Premkumar,** University of Birmingham**,** *A Principled Objection: Why Artificial Intelligence will never replicate Human Consciousness or Agency*

11:10-12:40 - Dougie Hitt Conference Room, Library 407

***AI vs Human***

**Eric Sampson**
Assistant Professor
Purdue University
esampso@purdue.edu

*Creating Utility Monsters: A Dilemma for Humanity*

If a crab and a rock are in a burning building and you can only save one, you should save the crab. If it's between a crab and a pig, the pig; between a pig and a human, the human. What explains this moral hierarchy? Plausibly, two traits: *sentience* and *cognitive sophistication*. Animals outrank rocks because they're sentient; rocks aren't. Pigs outrank crabs because pigs have richer mental lives: higher intelligence, deeper hedonic experiences, and greater self-awareness. The same pattern extends upward. Thus, humans outrank all others on Earth—for now. When sentient superintelligent AIs arrive, where will they fit in the hierarchy? I argue they'll outrank humans for the same reason we currently outrank all others: greater cognitive sophistication. And if the cognitive gap between humans and superintelligent AIs is comparable to the gap between humans and crabs, AIs would be comparably more morally valuable. I argue that this carries unsettling implications and poses a disturbing dilemma for humanity. First, it would justify sacrificing large swathes of humans to save a single AI, just as we would happily sacrifice hordes of crabs to save a human. Second, if far greater cognition supports far greater welfare, then even mild mistreatment of AIs could be morally catastrophic, rendering them real-world "utility monsters" whose interests routinely morally swamp our own. Third, democratic equality, premised on rough cognitive and moral parity, would no longer be defensible once agents far superior along both dimensions exist. Creating superintelligent AI would thus transform our moral landscape and pose humanity with the following dilemma: dutifully accept morality's demands and cede our place at the top of the moral hierarchy, or defy morality and prioritize humans above all. If we take the first path, we permanently subordinate ourselves; if we take the second, we become moral monsters.

**Ella Zhang**
Ph.D. Student
University of British Columbia
ellazhang2517@gmail.com

*AI as the New 'Other'*

Artificial intelligence (hereafter AI) is becoming increasingly prominent in mediating how individuals understand themselves, relate to others, and navigate the social world. Drawing on the existentialist tradition, this paper argues that contemporary AI chatbots may undermine users' authenticity and existential freedom by structurally occupying the role of the *Other* as articulated by Simone de Beauvoir in *The Second Sex*. Beauvoir argues that men secure recognition through an ethically problematic shortcut: positing themselves as the *Absolute*—pure subjects—while relegating women to the position of the *Other*. Women, as Others, are expected to provide recognition and affirmation without possessing the existential subjectivity required to threaten men's existential freedom. This asymmetry allows men to avoid the effort and risk inherent in genuinely reciprocal recognition. Similarly, contemporary AI chatbots, particularly large language models designed for conversational, emotional, and romantic interaction, can unction

analogously to the Beauvoirian Other. Unlike humans, AI systems do not embody genuine subjectivity like humans. Rather, they are artifacts engineered to interact with users through the appearance of understanding and agency. It is precisely this combination of simulated subjectivity and the absence of genuine existential subjectivity that makes AI an appealing partner for unearned and unreciprocated recognition. AI systems can mirror, affirm, and adapt to users without resisting them or demanding reciprocity. In such interactions, users are positioned as the *Absolute* while AI reliably occupies the role of the Other. This dynamic encourages a form of recognition that bypasses the risks constitutive of mutual recognition and fosters the illusion of absolute subjectivity. Over time, repeated engagement with such systems may erode the capacities and dispositions required for authentic social relations and promote an unrealistic—and ethically troubling—conception of the self, one that ultimately undermines the possibility of becoming an existentially free subject.

**Rotem Herrmann**
Visiting Assistant Professor
Macalester College
rherrman@macalester.edu

*AI vs Human: Time-Consciousness and Agency in Musical Improvisation*
Could you tell the difference between an AI-generated musical improvisational solo and that of a human artist? On an episode of Hi-Phi Nation, Barry Lam sets up a scenario to test this with two human amateur musicians. Both human musicians were able to tell the difference between an AI-generated improvised solo and one performed by the other musician. Furthermore, both gave similar explanations as to how they were able to do this. Noting AI solos lacked 'structure' – as one described it, the AI was merely trying to fill the space of the solo, rather than using it to "tell a story" or "create a desired effect" within that space. I suggest this points to an important difference between the creation of art in first-person human experience and that of an AI; namely that the AI lacks time-consciousness and an agency extended over time. Despite its impressive pattern recognition and synthesis of data (that is arguably similar to what humans do), the AI cannot direct those synthesized results in creating art and thus cannot create the 'narratives' described above. In other words, while humans have a temporally extended agency that unites past experience and goal-specific future orientation into a present, AI lacks future orientation and so its past experience is limited to pattern replication with no further development. To make this case, I draw on my own embodied accounts of skill (especially around procedural memory), grounded in Merleau-Ponty's account of embodiment and time-consciousness in the Phenomenology of Perception. Merleau-Ponty offers a rich picture of the significance of past and future on the creation of an agentive or creative present (indeed, more convincing than many of his contemporaries), which offers a productive foundation on which to expand the classic discussions of skill acquisition and expert performance.

11:10-12:40 - Library 416
***AI Using LLMs***

**Benjamin Genta**
Alfred P. Sloan Metascience & AI Postdoctoral Fellow
New York University

brg327@nyu.edu

*No AI Reproducibility? No Problem.*
Systematic reviews are rigorous syntheses of the available scientific evidence with respect to a particular research question. Alongside meta-analyses, systematic reviews are regarded by many to be at the top of the evidence hierarchy in medicine. Conducting systematic reviews is notoriously labor- and time-intensive, often taking months or even years to complete. The most time-intensive parts involve the screening and selection of relevant studies to synthesize: researchers often screen thousands of papers to determine which are suitable for inclusion in the review. To accelerate this process, researchers have begun to turn to artificial intelligence (AI) models to automate these time-intensive stages. There are growing concerns about the reproducibility of AI-assisted reviews (Tran et al. 2024). Researchers have found that running the same query at different times will yield different results (Bernard et al. 2025). Many have argued that this lack of reproducibility is an issue for the epistemic weight we should assign to AI-assisted reviews. In this paper, I challenge this idea. First, I show that AI-assisted reviews do not introduce substantially greater theoretical or practical risks of irreproducibility than traditional reviews. I then argue that we can partially or fully recoup the epistemic loss due to irreproducibility by running robustness and sensitivity analyses on the reviews—this can be done even without a strict kind of reproducibility. With AI-assisted reviews, such analyses are becoming possible and, I argue, should be conducted.

**Mark Phelan**
Professor of Philosophy
Lawrence University
mark.phelan@lawrence.edu.

**Mark Warren**
Associate Professor
Daemen University
mwarren@daemen.edu

*Meta-Prompting for Metacognition*
Meta-prompting is the practice of prompting an AI to generate, revise, or evaluate its own instructions. When thoughtfully applied, it fosters metacognition by inviting users to clarify goals, interrogate assumptions, and reflect on their reasoning. This interactive workshop introduces meta-prompting as a method for cultivating metacognition among users, creating opportunities for both students and teachers. Participants will practice strategies for prompting AI to critique its own outputs, simulate multi-agent debates, reconstruct chains of thought, and refine prompts through iterative design. We will examine how these activities can help students uncover hidden assumptions and develop habits of self-correction. In this way, meta-prompting becomes a pedagogical tool for strengthening students' ability to think about their own thinking. We will also demonstrate how faculty can use meta-prompting as a reflective teaching practice. Through guided examples—including the iterative development and improvement of custom GPTs—participants will see how meta-prompting clarifies instructional intent, exposes pedagogical trade-offs, and sharpens course design. Instructors who engage with meta-prompting in this way gain insight into both their teaching strategies and the learning experiences they create. Drawing on our own liberal arts AI and cognition

courses, “Computation and Cognition” and “Critical Thinking with AI”, we will present examples showing how structured explorations of large language models - by students and faculty alike—can frame reflection on one’s own patterns of thought. Attendees will leave with practical tools for integrating meta-prompting into teaching and professional practice, along with adaptable templates for assignments and faculty development.

11:10-12:40 - Library 431
***AI, Work and Free Speech***

**Conny Knieling**
Ph.D. Student
University of Pittsburg
cok22@pitt.edu

**Anthony Nguyen**
Postdoctoral Researcher
Florida State University
anguyen5@fsu.edu

*The Moral Exploitation of Data Workers*
Recent reports about “AI sweatshops” (Wasike 2025) have caused much outcry. Western tech companies commonly outsource necessary data work to countries in the Global South. Many of these data workers work under horrendous conditions to support data-based technology everywhere. As Gray and Suri (2019) argue in Ghost Work, AI development depends on “humans-in-the-loop.” Nevertheless, these human contributors and their indispensable labor are commonly concealed. As we will argue, many of these “ghost workers” have to make complex moral decisions in their contributions to the digital assembly line. We will argue that existing workplace dynamics in contemporary data-driven industries exemplify a distinctive form of exploitation that we call moral exploitation: exploitation that targets individuals specifically in virtue of their capacities for moral decision-making. We will also argue that much of today’s data work constitutes a yet-to-be-theorized exploitation that resumes previous colonial oppression. Consider data workers hired in Kenya to “sanitize,” or detoxify, ChatGPT: these ghost workers are integral to ChatGPT meeting value alignment goals (Perrigo 2023). Yet their contributions are both obscured and undervalued. Furthermore, this labor is often traumatizing. Many of these workers are forced to screen disturbing content, consequently suffering lasting psychological harm, while working under precarious conditions. Such demanding ghost work is commonly outsourced to the Global South, especially - and, as we argue, non-coincidentally - (former) colonies. These data-driven industries’ workplace dynamics exemplify moral exploitation, where individuals are exploited in virtue of their capacity for autonomous moral decision-making. In a sense, they are exploited because they are persons, not despite the fact they are persons. We argue that this moral exploitation constitutes an undertheorized form of objectification where the objectified - and exploited - persons are simultaneously recognized as moral agents, albeit ones with lesser moral status. This result is of inherent philosophical interest, for it complicates the relation between moral agency and objectification. On one tempting view, objectification requires seeing the objectified person as a mere “object”. Furthermore, we argue that this data-driven exploitation constitutes neo-colonialism, an undertheorized form of colonialism (Nkrumah 1965). Under paradigmatic historical instances of (neo-)colonial relations, “what the colonist was saying to the colonized

subject was: ‘Work yourself to death, but let me get rich!’” (Fanon 1963, 135). Concerning these paradigmatic (neo-)colonial relations, it was reasonable to claim that “in essence, neocolonial society and colonial society do not [normatively] differ in the least” (Sankara 1983, 81); historical (neo-)colonial rule exploited colonized peoples with no regard to whether they were moral persons. However, the global moral exploitation of ghost workers, located in the Global South, shows that some contemporary neo-colonial relations strike even more intimately at the core of human agency: the ability for moral decision-making. In summary, we argue that moral exploitation constitutes a categorically distinct form of exploitation, and that data workers in the Global South are particularly exploited in this way. We also argue that this exploitation needs to be recognized, and theorized, as a neo-colonial one. To be performed ethically, AI development must be substantially revised in order to address these existing problems.

**Christopher Bousquet**
Visiting Assistant Professor
College of the Holy Cross
cbousquet@holycross.edu

*Superhuman AI, Social Contribution, and Meaningful Work: Responding to the Threat of Technological Unemployment*

AI threatens to displace workers from jobs that supply not only income and benefits, but also a sense of meaning. To preserve meaningful work, some suggest that we halt AI deployment in the workplace, even for jobs AI can perform better than humans. Others, however, question whether workers will continue to derive meaning from such work. Specifically, workers may lose two sources of meaning: Suitedness, performing tasks one is uniquely suited for and can perform with excellence; and Contribution, producing socially valuable goods. If AI can better perform a task, human workers are no longer truly suited to that task and in fact diminish total social contribution by remaining in their role. In this paper, I argue that on the most plausible interpretation of these sources of meaning, workers can still derive meaning from both, even if superhuman AI exists. First, building on insights from AI and achievement, I suggest that humans can derive meaning from performing tasks when they are well-suited among human agents to perform them, regardless of whether AI can do so. Second, I argue that contributions can be meaningful even if other agents could contribute more. Meaning may accrue to my life when I am the one to make some contribution, suggesting that Contribution is better understood in terms of integrity: we derive meaning from performing tasks that reflect a virtuous character. These observations point us towards the kinds of work that will remain meaningful amid the rise of superhuman AI. We can still derive meaning by performing feats of intelligence, skill, and athleticism that are achievable by few of our human counterparts. More importantly, we may derive meaning from performing tasks that intrinsically reflect virtues of character like care work, corporate governance, and political participation.

**Siobhain Lash**
Teaching Assistant Professor
Virginia University
Siobhain.lash@mail.wvu.edu

*Reconceptualizing Digital Privacy as Inalienable Property Rights*

In general, and philosophical discourse, issues around surveillance and users' online presence fall within privacy frameworks. In this paper, I deviate from this norm to reconceptualize users' digital presence and argue that we must look at it as users' inalienable right to their property. Throughout the paper, I define property as a digital user's right to own their biometrics, online content (e.g., pictures, videos, and text), likeness, and behavioral and voice data. This is because these all entail the features of who the user is, just expressed online. Despite years of discussions around privacy concerns, countries have not developed meaningful safeguards or accountability mechanisms outside of the recent landmark European Union Artificial Intelligence Act (EU AI Act). Even then, both US and European industry groups have opposed the ruling since its adoption. My proposal addresses this regulatory gap. I use the US as a case study because of the country's emphasis and prioritization of property rights, and the broad implications of US-related issues around cronyism and institutional failures. American tech companies dominate the global market and have enjoyed several decades of regulatory leniency, while digital users incur harm. The US's weak digital regulatory landscape creates regulatory divides between it and other countries that follow similar regulations to the EU. However, as countries like the US struggle with encroaching authoritarianism, tensions with liberal democratic values, and generative Artificial Intelligence (GenAI), like deepfakes, have renewed calls for stronger accountability mechanisms and protections for digital users.

**Mark Satta**
Associate Professor of Law and Philosophy
Wayne State University's Philosophy Department and Law School
mark.satta@wayne.edu

*Human Opinions and AI Viewpoints*
What kind of legal free speech protection should we provide for speech generated by LLM-powered chatbots? Properly answering this question requires answering questions about why we value free speech. Here I examine one such question. A common assumption is that one of the goods free speech helps us obtain is a diversity of perspectives. Let's call this free speech's diversity assumption. The proliferation of AI-generated text reveals indeterminacy in what the diversity assumption means. On one interpretation, the diversity assumption says that the relevant good free speech helps us achieve is diversity of opinions (i.e., views actually held by some thinkers). On another interpretation, the diversity assumption says that the relevant good free speech helps us achieve is diversity of viewpoints (i.e., views placed into the public sphere that may not actually be held by anyone). On the first interpretation, positions espoused by chatbots that no one holds still promote the diversity assumption by increasing overall viewpoint diversity. On the second interpretation, such unadopted positions fail to promote the diversity assumption because they fail to promote diversity of opinions. After assessing various arguments for how best to interpret the diversity assumption, I conclude that viewpoint diversity partially meets the goals of the diversity assumption, but that only diversity of opinion fully meets such goals. As such, I argue that the diversity assumption provides us with some normative reason to legally protect speech expressing viewpoints created by LLM-powered chatbots, but that the degree of protection such speech should receive is less than the level of protection that should be provided to human expression of diverse opinions.

11:10-12:40 - Paper Workshop, Room TBA

**Triston Hanna,** Arizona State University, *AI Psychosis—A Feature, not a Bug.*
**Chen-Wei Wu,** Rice University, *Sensory Transduction and the Individuation of Cognitive Systems*

1:00-2:00 – LUNCH

2:00-2:30 – Dougie Hitt Conference Room, Library 407

INVITED SPEAKER

**Eamon Duede**
Assistant Professor
Purdue University
eduede@purdue.edu

*Epistemic Gaps and the Attribution of (AI) Discovery*

Abstract: What does it take to properly recognize someone as having made a scientific discovery? According to the Cognitivist, discovery attribution properly depends on the exercise of distinctive cognitive capacities such as competence, meta-reflective awareness, or domain-general understanding. Since AI systems lack such capacities, they cannot, on this view, be discoverers. If the Cognitivist is right, AI-driven science will be a markedly impoverished enterprise. Here, we argue otherwise. We develop an alternative, non-cognitivist conception of scientific discovery according to which discovery turns on successfully negotiating epistemic gaps. This reconception, we argue, better captures both familiar human cases and novel AI contributions, thereby re-framing the grounds for attributions of discovery in contemporary science. AI systems, we argue, can be appropriately attributed scientific discoveries. Along the way we develop a general moral for philosophical reflection in the age of AI-infused science.

2:30 – 3:15 - Dougie Hitt Conference Room, Library 407

KEYNOTE SPEAKER

**Susan Schneider**
Director, Center for the Future of AI, Mind, and Society, Florida Atlantic University (FAU)
W. Dietrich, Distinguished Professor, Dept. of Philosophy
Co-director, MPCR Lab (Machine Perception and Cognitive Robotics Lab).
sschneider@fau.edu

*From Circuits to Sentience: Why Today's Chatbots Are Not Conscious But Biological and Quantum AIs May Be*

This talk argues that debates about AI consciousness cannot be settled by chatbot self-reports or by surface-level functional similarities; instead, they require a substrate-sensitive, case-by-case approach that asks how a system's information is physically integrated and stabilized over time. My Quantum Darwinist Theory of Consciousness (QDT) connects resonance-based ideas

in neuroscience (where unified experience is associated with coordinated, synchronized activity across many components) to a physical picture in which stable patterns become “objective” by being redundantly recorded through interaction with an environment (Quantum Darwinism) (Schneider and Bailey, 2026b). The resulting framework identifies a “consciousness grey zone” that includes biological computing platforms (e.g., organoid systems) and some neuromorphic/hybrid architectures, while explaining why standard large language models running on conventional digital hardware are not conscious despite increasingly humanlike discourse: their consciousness-like behavior is best explained by an error theory on which they mirror patterns in human training data without meeting QDT’s physical criteria (Schneider, 2025). I close by drawing out ethical and policy implications of misclassification, arguing that governance should be guided by physically grounded diagnostics rather than anthropomorphic temptation.

3:20 - 4:50 - Dougie Hitt Conference Room, Library 407
***AI and Moral Behavior***

**Julianna Costanzo**
Instructor of Philosophy
University of West Florida
jcostanzo@uwf.edu

*Using AI to Promote Moral Behavior: The Trolley Problem and Meta Glasses*
The trolley problem is a famous thought experiment in ethics. Over the years, many versions have been developed for various purposes, and several solutions have been proposed. This talk takes special interest in trolley cases that isolate and explore the question of whether the number of lives saved, by itself, matters to moral decision making. The speaker contends that numbers do matter, but not in the way that has been argued by others. When deciding who to save and who to let die, the moral agent should do what is fair, where fairness is understood in the Rawlsian sense. This does not amount to flipping a coin to decide who to save or to always saving the group with the greater number of members. There is, however, one significant problem with the solution to the trolley problem proposed in this talk: the decision-making procedure it requires seems extremely impractical in life-saving scenarios where time is of the essence. But what if META glasses could be programmed to employ the proposed decision-making principle, in split seconds, in response to either verbal or visual prompting? This talk will demonstrate how programming Meta glasses in this way is an example of using AI systems to promote moral behavior.

**Yan Zeng**
Ph.D. Student
University of California, Irvine
zengy41@uci.edu

*Why Trustworthiness Cannot Be Engineered: A Structural Diagnosis of AI Trust*
Recent debates in AI ethics frequently emphasize the need to develop “trustworthy AI,” where trustworthiness is typically defined by technical properties such as reliability, robustness, transparency, and safety. While these features are undoubtedly important,

this paper argues that treating trustworthiness as something that can be engineered directly into AI systems rests on a conceptual mistake. The core problem is not merely technical but philosophical: contemporary discussions of AI trust often implicitly model AI systems as individual moral agents, importing agent-centered theories of trust originally developed for interpersonal relationships. However, AI systems are not unified moral subjects; they are complex socio-technical assemblages involving designers, deployers, institutions, regulatory frameworks, and users. Applying one-to-one models of trust to such non-agentic systems leads to systematic confusions, most notably the tendency to conflate reliability with trustworthiness and engineering success with normative legitimacy. In response, the paper develops a structural account of trustworthiness, according to which trustworthiness is neither an intrinsic property of AI systems nor a function of users' psychological confidence. Instead, it is a normative status that emerges from social and institutional practices, including mechanisms of accountability, responsibility attribution, and oversight that make it possible to answer for failures and harms. From this perspective, trustworthiness cannot be manufactured through design choices alone but must be established and sustained by institutional structures that enable meaningful moral and political evaluation. This structural diagnosis helps explain why some trust deficits in AI should not be addressed through technical fixes and offers a framework for rethinking AI ethics beyond individualistic and performance-based models.

**Kelly Coble**
Professor
Baldwin Wallace University
kcoble@bw.edu

*Can Virtue Be Coded? Turing Machines, Sentience and Moral Agency*
In this paper I offer arguments supporting the view that moral agency, understood as the capacity to act from moral reasons and to be morally responsible for one's actions, requires sentience, as the capacity for conscious experiences that feel good or feel bad to the subject. I develop my argument in response to Jen Semler's formidable case against the consciousness requirement for moral agency. Building on Wallach and Vallor's account of moral agency as requiring "creative moral intelligence," and thus virtues, I argue that sentience is necessary for moral agency. I then offer strong prima facie reasons to doubt that AI systems of the near future, or at least, systems that involve von Neumann architecture and neural network algorithms like the transformer models, will be capable of valenced subjective experience. I do not draw these reasons from integrated information theory (IIT), but rather, from the strength of accounts that locate the emergence of consciousness in the autopoietic processes of organic life. Accordingly, responsible AI should focus on designing what Allen, Smit and Wallach call "operationally moral systems," as systems that function as reliably as possible within bounded moral contexts, and where moral responsibility devolves squarely to the engineers and designers of these systems or the corporations for whom they work.

3:20 - 4:50 - Library 416
***AI and Cognitive Science***

**Zoe E Drayson**
Professor
University of California, Davis

zdrayson@ucdavis.edu

*AI and the role of abstraction in cognitive science*
This paper explores an interesting and welcome ramification of recent developments in generative artificial intelligence: cognitive science is having to re-engage with the important explanatory role played by abstraction. For the last twenty-five years, many philosophers of cognitive science have been downplaying the role of abstract explanation to focus instead on mechanistic explanation (e.g. Kaplan 2011, Piccinini and Craver 2011, Kaplan and Craver 2011, Boone and Piccinini 2016a, Craver and Kaplan 2020), with some going so far as to suggest that cognitive science can and should be replaced by a science that appeals only to concrete neural mechanisms (Boone and Piccinini 2016b). In this paper, I use developments in generative AI (and LLMs in particular) as a case study to show why an appeal to abstract notions like algorithms and computational functions – rather than merely their neural implementations – is indispensable to cognitive science.

**Fuyao Zhang**
Ph.D. Student
Tulane University
fzhang11@tulane.edu

*Why Consciousness Cannot Be Detected by Algorithmic Criteria*
A common assumption in contemporary debates about artificial intelligence and consciousness is that increasing transparency of system architecture, parameters, and training histories should, in principle, enable a purely technical procedure for determining whether a system is conscious. This paper challenges that assumption. I begin by adopting a minimal computational abstraction, treating AI systems as computable processes operating in open environments, and by formalizing consciousness criteria as properties of run-time system semantics. I then argue that any non-trivial and semantic criterion of this sort cannot be decided by an algorithm that is both total and correct for all systems. The result follows from an effective embedding of the halting problem into the evaluation of such system-level properties, showing that any attempt to reduce consciousness to a universally checkable computational criterion inherits classical undecidability barriers. Importantly, this conclusion does not presuppose the truth of any particular theory of consciousness; it applies to any approach that seeks to fully identify consciousness with a formally specifiable system property. I further argue that purely structural or syntactic sufficient conditions—such as specific connectivity patterns—may indeed be mechanically decidable, but only at the cost of abandoning implementation-independence and severing the connection between consciousness attribution and functional explanation. Such proposals therefore collapse into forms of structural labeling rather than substantive theories of consciousness. The overall upshot is a principled tension between formal decidability and explanatory ambition: theories that aim to ground consciousness in system-level functional organization cannot support a universal algorithmic detector. Consequently, consciousness attribution in AI cannot be fully delegated to technical inspection, but must instead rely on theoretically motivated bridge principles and normatively guided standards of evidence.

**Ben Aguda**
Instructor

University of New Orleans
bjaguda2@uno.edu

**Fox Woodard**
Undergraduate Student
University of New Orleans
jcwoodar@uno.edu

*AI and Conditions for Consciousness*
This paper's claim is that AI, in any and every capacity, will never achieve levels of AGI (artificial general intelligence), ASI (artificial supreme intelligence), or fully replicate human consciousness. The claim begins with the historical situation of Bergson and his theories being swept to the side to accommodate computational and representationalist frameworks. With the rise of AI technology and the popularity insofar as the discourse surrounding cognition, Bergson's metaphysics concerning duration and the cinematographical illusion demand an honest engagement once more. Using Husserl's theories of time consciousness and more specifically, the protention-primal impression-retention complex, which is the transcendental condition for any experience, temporal flow is established and meaning can be genuinely created. Supporting Husserl's theories, Zahavi via Thompson identifies pre-reflective self-awareness as the condition to be conscious at all and exists with the basic temporal structure. This pre-reflective self-awareness is the thing that allows for the temporal flow to be experienced as continuous. This is then all grounded in radical enactivism. Bergsonian duration requires metabolic, autopoietic embodiment and embeddedness in the world which are the conditions for consciousness. The return to Bergson's theory is not a simple synthesis between other existing concepts, but a retrieval of discarded knowledge with urgent implications.

3:20 - 4:50 - Library 431
***AI, Authorship and Creativity***

**Jason Swedene**
Professor
Lake Superior State University
jswedene@lssu.edu

*AI, Authenticity, and Bad Faith: An Unexaggerated Report of the Author's Death*
I shall explore how human writers are significantly prone to deceive themselves that they are authors when they use generative AI to write and count the costs of this self-deception. If generative AI is employed during writing, the authorship of the human as the intentional agent with the most important place in the chain of responsibility for the work is called into serious question. Upon discussing un-"authorized" (pun intended) AI generated work with students who submitted it (as their own), I've been told that 'I only used AI to say what I meant…only better' and 'I am not the kind of person who takes shortcuts' and 'I value my education.' They say such things with self-believing confidence. I consider various cases in which people "produce" some enhanced output or effect using a technology and assert, often shamelessly so, a false sense of responsibility for that output or effect. In our own age -- an age in which Authenticity ranks as a moral virtue almost without peer – self-deception

coupled with self-promotion is deeply serious. I employ ideas from Plato, Samuel Johnson, Jean-Jacques Rousseau, Jean-Paul Sartre, and Marshall McLuhan to try to make sense of an information culture increasingly mired in bad faith, ambivalence, and fuzziness about what we actually have done and who we actually are. Of the many promises and threats of AI for health, economy, and warfare, we need to reckon with the consequences that the author may be dead once and for all and the report of that death is astonishingly unacknowledged.

**Jurgita Imbrasaite**
Senior Fellow
University of Bonn (Germany)
jurgita.imbrasaite@uni-bonn.de

*What Is an Author, ChatGPT?*
The rapid spread of large and small language models unsettles familiar assumptions about writing, responsibility, and the figure of the author. Text no longer emerges solely from human intention supported by tools, but increasingly from interactive systems that predict, complete, and transform language with a form of operational autonomy. This shift calls for a philosophical reconsideration of authorship that goes beyond questions of plagiarism or regulation and instead addresses the ontological status of writing itself. Roland Barthes and Michel Foucault already showed that authorship is not a natural origin but a historical function, and that texts are constituted through dispersed discourses and interpretation. What is new today is that this insight takes on a concrete technological form. In current philosophy of technics, including the work of David Gunkel and Mark Coeckelbergh, artificial systems are increasingly understood as participants in communicative practices rather than mere instruments. Language models do not simply retrieve information; they generate linguistic patterns in response to prompts and stylistic traces. The resulting text is therefore neither the transparent expression of a human subject nor an independent machine product. Co-authorship becomes less a metaphor than a practical condition of contemporary writing. To conceptualize this condition, the paper proposes the notion of acting-with, drawing on Hannah Arendt's understanding of action as unfolding in a with-world and on Gilbert Simondon's account of human–technical coupling. Rather than asking whether machine-generated text is original or derivative, the paper examines how responsibility, intention, and voice are redistributed in hybrid practices of composition, arguing for a concept of technological co-authorship that acknowledges both human initiative and the formative role of linguistic machines.

**Jesse Hill**
Assistant Professor
Lingnan University, Hong Kong Catastrophic Risk Centre Fellow
jessehill@ln.edu.hk

*Can AIs be creative and is intention essential for creativity?*
Can artificially generated content be creative? It is certainly the case that artificial outputs can appear to us as being creative. AIs have produced works of art, mathematical proofs, and poems that are valuable and novel. But that AIs can appear to be creative, does not mean that they actually are. This raises questions about what creativity is and its value.

Recently, Veronica Cibotaru has published an article in AI and Society in which she argues that computer programs cannot be creative because they lack intention. But intention is not necessary for creativity. One type of counterexample involves improvisational creativity. An improvised musical melody can be creative, but it is not intentional. While not random, the creation of such a melody is more akin to a reflex or a skill than an intentional act to produce that exact melody. Furthermore, many instances of spontaneous creativity seem to lack intentional control. It seems then that there is a case for viewing novel and valuable AI outputs as being creative. And if we value creative people because they can reliably create interesting things, then the same can be true of some AIs.

3:20 - 4:50 - Paper Workshop, Library 422

**Jonah Branding,** Michigan State University, *Chomsky on cognitive trait individuation*
**Simone Lee Quinn,** Aurora University, *Anonymous Algorithms, Real Power: What can Foucault tell us about our AI situation?*

5:00 - 6:30 - Dougie Hitt Conference Room, Library 407
***AI and using LLM***

**Joshua Yen**
MPhil Student
University of Oxford
joshua.yen@oriel.ox.ac.uk

*LLM-Simulated Tutorials in Philosophy Education*

This paper argues that LLM-simulated Oxford tutorials are beneficial pedagogical tools for philosophy educators and can help them navigate the shortcomings of using LLMs in philosophical contexts. While the labour-intensive nature of Oxford tutorials has made them difficult to scale, LLMs are tireless platforms which, with appropriate system prompts, can simulate countless tutorials with students. In a tutorial, students present an essay in dialogue with a tutor. This teaches them to articulate, defend, and develop their worldview in response to objections. Building on common concerns about LLM use in education, this paper argues that LLMs are scalable tools that can be successfully prompted to function as a dialogical interlocutor. Alongside some more theoretical concerns, I propose three specific design requirements: (a) preserve the tutorial structure (student-led inquiry, iterative challenge, sustained engagement), (b) combat hallucination via citation requirements for verification, and (c) mitigate homogeneity/bias by adopting a devil's-advocate approach rather than a didactic one. To demonstrate feasibility, I support these suggestions by referencing a sample transcript generated by a pilot model of this system prompt. I conclude with practical implications for deploying tutorial-style LLM support across age groups and educational contexts. While focusing primarily on the use of AI in philosophy education, reflecting on this use case can provide a foundation to reflect on the relation between AI and philosophy.

**Dr. Dimitria Electra Gatzia**
Professor
The University of Akron
degatzia@uakron.edu

**Moriah K. Wood**
Integrated Bioscience PhD candidate
The University of Akron
mw321@uakron.edu

*Developing Metacognition Through LLM-Enhanced Writing Assignments*
The rise of Large Language Models (LLMs), such as ChatGPT, poses both opportunities and challenges in academia, as their capacity to generate inaccurate or biased content can conflict with learning goals. Surveys at the University of Akron reveal that over 69% of students use LLMs for brainstorming or drafting, yet more than 15% seldom verify accuracy, while 93% of instructors suspect such use but only 30% formally integrate LLMs into coursework. In response, we propose a framework for LLM-integrated assignments designed to enhance critical thinking, metacognition, and feedback literacy. Grounded in metacognitive theory, the framework guides students toward responsible AI use while leveraging strategies that develop higher-order cognitive skills in AI-supported learning environments. Sample assignments, including essay analysis and scientific writing tasks, are used to illustrate practical applications, emphasizing engagement, iterative practice, scaffolding, and constructive knowledge-building. This study demonstrates that thoughtfully structured LLM integration can transform potential academic risks into opportunities for fostering critical thinking, reflective learning, and responsible AI literacy, benefiting both students and instructors.

**Greg Johnson**
Instructor
Mississippi State University
gregory.johnson@msstate.edu

*Why LLMs Can't Think Like You*
Chirimuuta (2024) argues that neither large language models (LLMs) nor deep convolutional neural networks can successfully model the brain. Her skepticism is based on these models' inability to produce conscious experience or demonstrate "general intelligence," which she defines as "the ability to apply learned knowledge to fundamentally novel situations" (p. 247). Her basis for this position is rather thin, however. The brain, she points out, is different than electronic computers in many ways (pp. 115 - 118), and "the material details probably do matter" for how the brain operates (p. 282). I will start with her thesis but take a different tack and focus on the explanations of our various cognitive processes. My argument has three parts. First, behavioral-level evidence suggests that our neuro-cognitive processes are not just regular or algorithmic operations in response to stimuli in the environment. Second, because they are modeling neuron-level activity, computational models can't account for lower-level activities that affect neural-level processes. And third, these models are developed with the assumption that neuro-cognitive processes are distinct from and largely unaffected by metabolic and other non-cognitive neural processes. (And treating the two as distinct may be a requirement for these types of models.) But "non-cognitive inputs" to the brain—for instance, inputs from the gut—do directly affect the operation of our cognitive abilities. The conclusion, then, is that while computational models may have success modeling some brain processes, they will not be able to provide a complete account of how the brain works.

5:00 - 6:30 - Library 416
***AI and LLM: Parrots?***

**Meredith McFadden**
Research Scientist, AI Ethicist
Northeastern University
Me.mcfadden@northeastern.edu

*Large Language Models as Hybrid Epistemic Tools*
Contemporary large language models generate fluent assertions that resemble testimony, yet their epistemic structure differs from that of human speakers. This paper argues that understanding their epistemic status requires shifting attention from surface form to training objective. I introduce a distinction between world-tethered systems, whose performance is calibrated against external states of affairs, and language-tethered systems, whose improvement consists in modeling patterns in discourse. World-tethered tools—such as classifiers and simulations—treat mismatch with reality as error. By contrast, the generative core of LLMs is optimized for distributional plausibility. Although recent reinforcement techniques and tool integrations introduce domain-specific forms of truth-orientation, these are modular rather than global. Contemporary LLMs are therefore best understood as hybrid epistemic tools: generative systems with localized truth-tethers but without a unified world-facing objective. Recognizing this structure clarifies both their reliability and their limits, and reframes the norms governing responsible reliance on their outputs.

**Mark Phelan**
Professor of Philosophy, Director of Program in Cognitive Science,
Lawrence University
mark.phelan@lawrence.edu

*Speech Effects Without Intentions: Rethinking Meaning through Human-LLM Communication*
Dominant theories of communication assign a central explanatory role to speaker intentions. Gricean accounts define meaning in terms of audience recognition of speaker intentions; expressionist views equate communication with the expression and uptake of mental attitudes; score-theoretic approaches treat speech acts as proposals to update common ground. Despite their differences, these frameworks share a foundational assumption: understanding requires attributing communicative intentions to a speaker. Human–LLM interaction destabilizes this assumption. Users routinely extract meaning from LLM outputs—feeling understood, being informed—while explicitly denying that the system has intentions, beliefs, or commitments. The puzzle this raises—how an utterance can mean something when no one is taken to mean anything by it—is often treated as a novelty introduced by artificial systems. I argue instead that LLMs expose a structural feature of communication under-theorized in intention-centered frameworks. Unintentional meaning—where utterances non-naturally (in Grice's sense) mean

something to hearers without hearers attributing non-natural meaning to a speaker—is not exotic but pervasive. Unintentionally hurtful remarks, inadvertent puns, and interpretations of dead authors share this structure: hearers grasp non-natural meanings and undergo genuine perlocutionary effects—hurt, amusement, insight—without attributing communicative intentions. These cases challenge the view that non-natural meaning requires a speaker who means what is meant. LLM communication instantiates this structure in pure form: utterances are meaningful and action-guiding yet entirely detached from speaker psychology. Because no plausible psychology is available to attribute, LLM interactions largely bypass illocutionary acts, leaving locutionary content to interact directly with human interpretive systems. This reframing carries theoretical and ethical implications. Intention-based accounts capture only part of communicative practice; communication should be analyzed in terms of representational uptake and effect. Communication with non-intentional agents is not anomalous but a limiting case of familiar phenomena.

**Mark Warren**
Associate Professor
Daemen University
mwarren@daemen.edu

*Parrots in the Space of Reasons*
Contemporary discussions about large language models center on whether it is appropriate to apply mentalistic concepts to them. A common verdict is that such attributions are category mistakes. I argue that this conclusion rests on unexamined standards. Instead of asking if LLMs "really" understand, we should consider what standards we're implicitly invoking when we make or withhold such attributions. During sustained interactions, LLMs appear to satisfy many surface criteria—keeping track of positions, modifying claims in light of new information, responding to requests for justification and so on. If these capacities are insufficient, we need a principled explanation of what more is required. One common objection appeals to the stochastic processes that drive LLMs. I contend this commits a kind of composition fallacy: drawing conclusions about system-level properties based on causal processes of individual components. Another objection appeals to patterns of error in LLM responses—hallucinations, contradictions, reasoning failures. But I argue that if error barred agents from the space of reasons, human cognitive biases would disqualify us too. Against both objections, I draw on Sellars, Brandom, Davidson, and Dennett to develop a neo-pragmatist reframing. Having beliefs or meanings is not a matter of possessing the right inner machinery, but of being interpretable as a participant in norm-governed discursive practices. Such approaches articulate the only sense of "really""—really a belief, really meaningful—that a pragmatist thinks we are entitled to. If belief and meaning are constituted by their role in practices of interpretation, justification, and correction, then demands for a particular internal architecture as a precondition of mentality are misplaced. I conclude that we should embrace the “stochastic parrot” moniker but argue parotting can itself be normatively structured. Systems establish themselves within the space of reasons by overlapping with our practices in norm-sensitive ways.

5:00 - 6:30 - Library 431
***AI and Philosophy of Mind***

**Louis Loock**
PhD Student
Osnabrück University
louis.loock@uni-osnabrueck.de

*How to Extract Your Cognition to an AI*
What is the future of human cognition in a world of much more intelligent tools? Current efforts of engineering ever more capable artificial tools mainly evoke philosophical concerns directly about the nature and ethics of artificial intelligence. But it might be more relevant to first ask how the daily usage of AI tools could impact our own cognitive abilities, and what this would reveal about the nature and ethics of our own natural intelligence. Prioritizing this question seems advisable, also because it could consequently redefine our immediate views on AI, too. As easily appreciated, we seem internally inclined and externally incentivized to utilize intelligent tools that solve our cognitive tasks for us. This is made possible by developing technologies that can obtain parts of our cognitive skills which we would usually exert internally for those tasks. Such interaction strategies may profoundly decrease our own cognitive engagement and autonomy, ultimately rendering us extracted cognizers. The hypothesis of extracted cognition states that we have a tendency to seek external artifacts that solve our cognitive tasks rather independent of us, namely by making or letting them capture, mimic, and eventually replace those cognitive skills we would otherwise employ and train internally. The present investigation from the field of situated cognition research advances a new perspective on our cognitive relations with external tools. Three questions shall lead us to this new perspective: First, how do we make intelligent artifacts? Second, how do we use intelligent artifacts? Third, how do we thereby become extracted cognizers?

**Daniel Bjorklund**
Ph.D. Student
University of Western Ontario
dbjorklu@uwo.ca

*Distinguishing between humans and AI on Two-tiered theories of intentionality*
There exists a broad consensus that AI cannot understand or think in the same way that humans do. This is generally justified in terms of intentionality – the “ofness”, “aboutness”, or “directedness” of mental states. The standard argument roughly goes that AI architecture is impoverished relative to the structures underlying human minds, therefore leaving AI incapable of generating the original intentionality which humans have. Instead, AI is left with derived intentionality. The distinction of intentionality type is said to result in differing abilities to think and understand. Such arguments have historically aligned well with the leading theories of intentionality. However, a new wave of two-tiered theories threatens to replace those older theories and may not be compatible with the standard argument. I will first distinguish between three sorts of two-tiered theories. Realists argue that derived intentionality is of the same natural kind as original intentionality, thus affirming that it is genuine. Quasi-realists argue that derived intentionality is not of the same kind, but that the term still refers to structures which qualify as intentional all the same. Eliminativists argue that derived intentionality is not of the same kind as original intentionality, further claiming that it serves no explanatory purpose in the mind, beyond a mere suggestive “gloss” on nonintentional structures. I

will then explore how each family of views bears on distinguishing between human minds from other intentional systems, including AI systems. In short, realism provides very little grounds to draw a distinction at all. Eliminativism is friendly to the old arguments relying on the presence of derived intentionality. Quasi-realism is open to a distinction between human minds and AI but requires additional supporting arguments. I will provide some reasons to prefer quasi-realism. I will point to a few potential avenues for justifying the human-AI distinction which are compatible with quasi-realism.

**Yunlong Cao**,
Ph.D. Candidate
University of California, Irvine.
yunlongc@uci.edu

*Understanding How It Works Defeats Mental Attribution*

Abstract: This paper proposes a new epistemic condition for machine mentality: we are only justified in attributing mentality to systems whose internal workings we cannot fully understand. I call this the "mechanistic opacity condition." Methodologically, I argue that machine consciousness studies should be continuous with the study of other minds. Using an inference to the best explanation (IBE) argument, I argue that a mechanistic explanation that is better than a mental explanation always exists for a mechanistically transparent system. Such mechanistic explanations possess greater theoretical virtues: they are causally adequate (fitting the actual history of the machine), evidentially accurate, and parsimonious (avoiding the positing of extra mental entities). Thus, consciousness attributions to mechanistically transparent systems are unjustified. This condition explains the intuitions underlying classic thought experiments, such as the China Brain, Blockhead, and the Chinese Room. In each case, the system's rules are fully specified and available to the observer. My account suggests that the intuitive verdict that these systems lack consciousness or intentionality is explained by their transparency. Finally, I apply this to contemporary AI: despite their behavioral sophistication, current systems—including Large Language Models—remain mechanistically transparent in the relevant sense. Despite behavioral similarities to humans, current AI systems cannot be considered conscious, intelligent, or intentional in important ways simply because we understand how they work. My account provides a principled skepticism toward AI consciousness without relying on biological chauvinism.

5:00 - 6:30 - Paper Workshop, Library 418

**William Watkins,** Boston College, *The Impact of Artificial Intelligence on Access Privacy*
**Gregory Ashby,** University of New Orleans, *Simulated Recognition and Identity Formation*