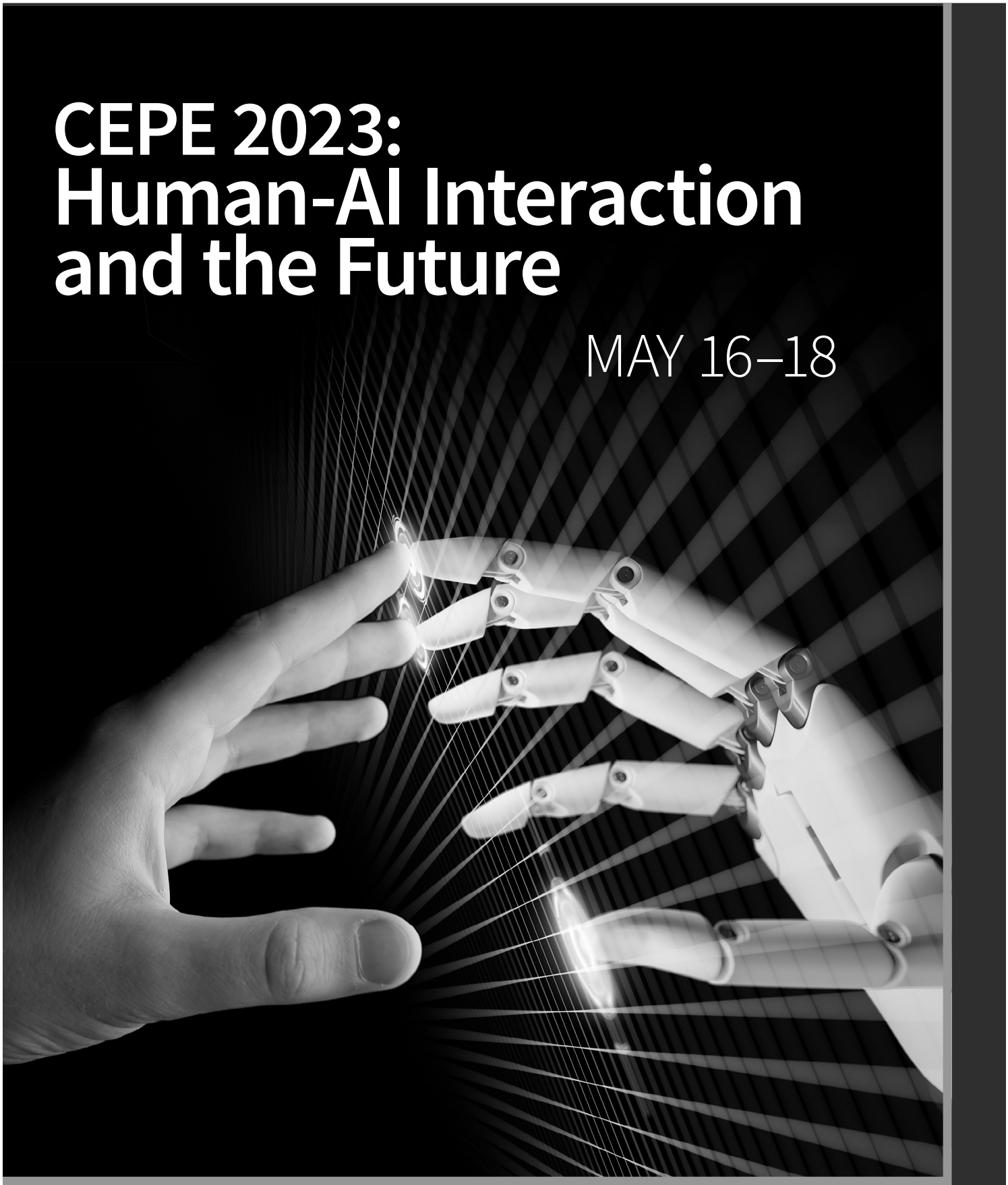
CEPE 2023: Human-AI Interaction and the Future

MAY 16–18

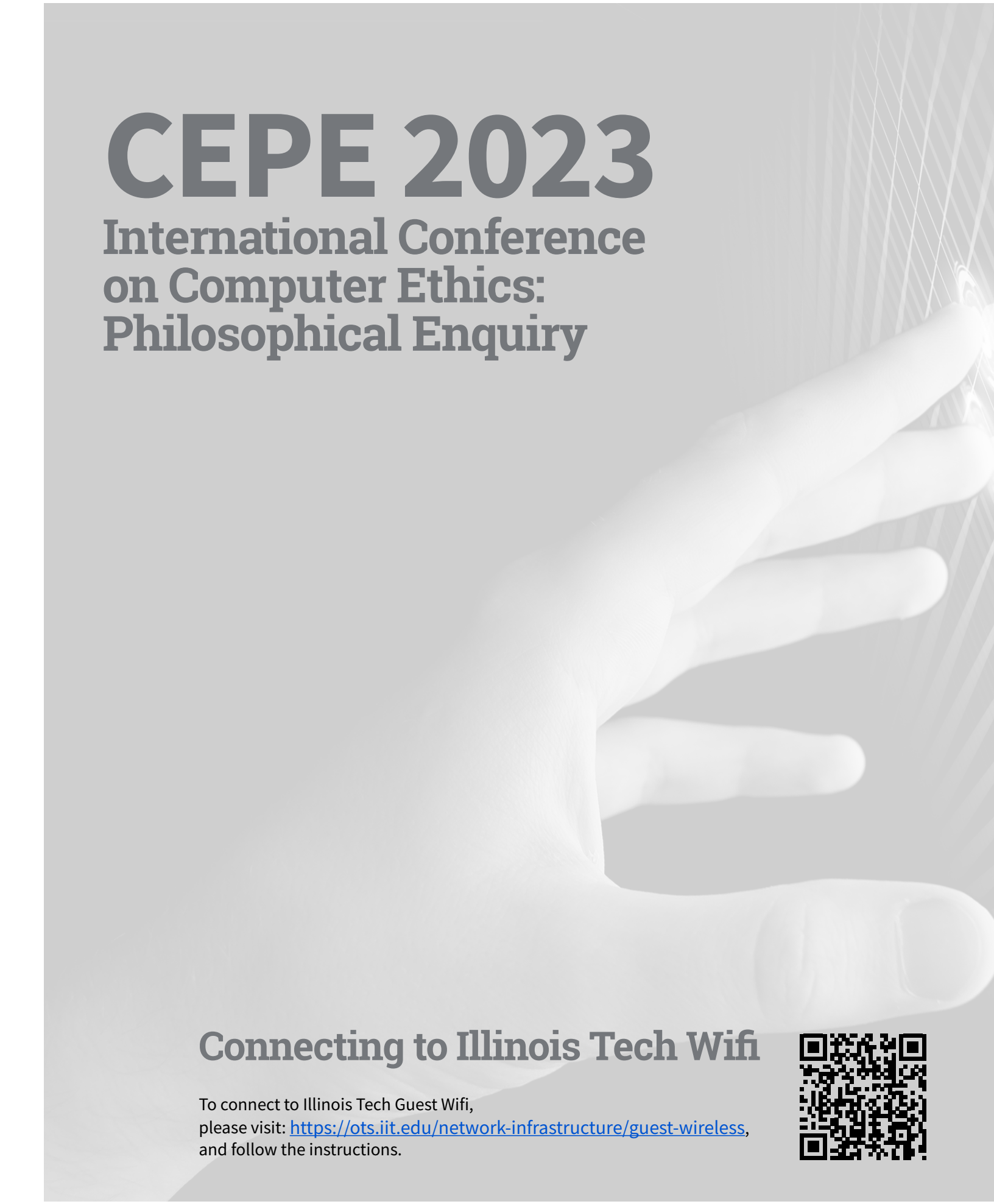


INSEIT
International Society for Ethics and Information Technology

ILLINOIS TECH | Center for the Study of Ethics in the Professions | Center for Cyber Security and Forensics Education

CEPE 2023

International Conference on Computer Ethics: Philosophical Enquiry



Connecting to Illinois Tech Wifi

To connect to Illinois Tech Guest Wifi,
please visit: https://ots.iit.edu/network-infrastructure/guest-wireless,
and follow the instructions.

Schedule at a Glance

Tuesday, May 16	Wednesday, May 17	Thursday, May 18
9:30–10 a.m. Coffee & Networking	**8:30–9 a.m.** Breakfast	**8:30–9 a.m.** Breakfast
10 a.m. Welcome	**9–10:30 a.m.** Parallel Session 3	**9–10:30 a.m.** Parallel Session 6
10:15 a.m.–12:15 p.m. Parallel Session 1	**10:30–10:45 a.m.** Break	**10:30–10:45 a.m.** Break
12:15–1:30 p.m. Lunch	**10:45 a.m.–12:15 p.m.** Plenary Session *Philip Brey*	**10:45 a.m.–12:15 p.m.** Parallel Session 7
1:30–3:00 p.m. Plenary Session *Helen Nissenbaum*	**12:15–1:30 p.m.** Lunch	**12:15–12:30 p.m.** Closing Session
3–3:15 p.m. Break	**1:30–3 p.m.** Parallel Session 4	
3:15–4:45 p.m. Parallel Session 2	**3–3:15 p.m.** Break	
5–6:30 p.m. Reception	**3:15–4:45 p.m.** Parallel Session 5	
	5–7 p.m. Dinner	

Day 1—May 16

10 a.m. **Welcome**

Elisabeth Hildt
Center for the Study of Ethics in the Professions, Illinois Institute of Technology

10:15 a.m.–12:15 p.m. **Session 1**

Room 1- Trust and Bias

Hermann Hall Ballroom
Chair: Elisabeth Hildt

When AI Moves Downstream
https://journals.library.iit.edu/index.php/CEPE2023/article/view/245
Frances Grodzinsky, *Sacred Heart University, United States*
Keith Miller, *University of Missouri, St. Louis, United States*
Marty J. Wolf, *Bemidji State University, United States*

In "On the Responsibility for Uses of Downstream Software" (2019) Wolf et al. explored the degree to and ways in which computing professionals are responsible for the downstream use of the software they develop; this analysis is based on the nature of the software itself, not on the nature of the downstream use. "Downstream use" refers to how a piece of software is used by others after its release.

The authors adapted a mechanism developed by Floridi (2016). In this work, Floridi shifted the question of responsibility away from the intentions of developers per se and onto the impact that their Distributed Moral Actions have on moral patients. Wolf et al. take this in a slightly different direction and make an argument that there are features of software that can be used as guides to better distinguish situations where a software developer might share in responsibility for the software's downstream use, from those in which the software developer likely does not share in that responsibility. The features of their Software Responsibility Attribution System (SRAS)—our term, not in the original paper —that they identified as significant are: closeness to the hardware, risk, sensitivity of data, degree of control over or knowledge of the future population of users, and the nature of the software (general vs. special purpose). In a subsequent paper, Grodzinsky et al. (2020), the same authors offered some evidence that these features and their impact on responsibility assessment are consistent with some sources in the literature.

Since that time, artificial intelligence (AI) has been increasingly deployed in many different application areas. In this paper, we will re-examine the SRAS with a focus using critical work on AI. That analysis will lead to adjustments to the SRAS. We will apply the modified SRAS to cases involving AI used in surveillance and AI used in social media.

Causes and Reasons – Decisions, Responsibility, and Trust in Techno-Social Interactions
https://journals.library.iit.edu/index.php/CEPE2023/article/view/270
Larissa Ullmann, *Technical University of Darmstadt, Germany*

The interaction between humans and AI creates a new type of interaction that goes beyond subject-object relations. AI technologies cannot always be described as a conventional object due to its autonomy and the black box aspect. An additional category is created, which is outlined by the 'sobject approach'. This creates the opportunity to study the human-like characteristics of the interaction on the part of the AI. The 'social' possibilities of AI can thus be focused by referring to 'techno-

social' rather than 'social' interactions, since the possibilities are different from the human sociality, but exist in the human-social lifeworld. If an AI is a techno-social interaction partner, it can 'act' and make 'decisions'. The additional category can therefore be used to investigate what types of decisions there are, if they are based on 'reasons or causes', whether they can be 'trusted', and if one can assign or delegate 'responsibility' to such technology. So, classical ethical questions regarding subjective categories like decision-making, trust and trustworthiness, and responsibility can be rethought for somewhat human-like but not human technologies like AI.

Overtrust in Algorithms – An Online Behavioral Study
https://journals.library.iit.edu/index.php/CEPE2023/article/view/285
Philipp Schreck, Artur Klingbeil and Cassandra Grützner,
Martin Luther University of Halle-Wittenberg, Germany

People inappropriately relying on technology can result in misuse of systems potentially leading to unethical decisions due to algorithmic bias, unprofitable outcomes for the affected parties or even safety hazards when operators fail to assume control in a situation where machines errs. While evidence hints on overtrust in algorithms, further research is required to understand which factors promote it and which contrarily foster distrust in algorithms. However, most of the research has focused on specific scenarios and limited tasks. Hence, studies on ethically relevant recommendations by algorithms are lacking a generalizable experimental setup that is context-independent of general factors that may lead to overtrust.

To address this, we utilize methodologies from behavioral economics to conduct online experiments. We assess under which conditions subjects demonstrate overtrust in counter-intuitive AI recommendations during decision-making situations and when subjects actively reverse the algorithm's recommendation. For this purpose, we manipulate factors such as available information about the algorithm, its perceived competence, reliability, explainability, rule-based vs. learning algorithms, decision complexity, decision amount and restrictions such as budget or time limits. The goal is to derive more generalizable principles from a behavioral perspective about how to design ethical AI that does not foster potentially harmful overtrust in machines.

Room 2 - Moral Decision-Making and AI - Societal Adaptation to AI

Hermann Hall Expo
Chair: Kelly Laas

The Overdemandingness of AI Ethics
https://journals.library.iit.edu/index.php/CEPE2023/article/view/241
Susan Dwyer, *University of Maryland, United States*

In unreflectively deploying moral concepts and principles 'designed' for human conduct, current AI ethics (ironically) holds AI to moral standards higher than those to which we hold each other in ordinary human life. Some might suggest that this is as it should be, since AI is, in a variety of ways, more powerful or impactful than ordinary human agency. But if this is so why haul in ethical concepts and principles designed for less powerful human beings? I argue that AI ethicists should neither abandon traditional ethical concepts or settle for a cynical 'ethics-lite' as business ethicists have. Rather, they should think systematically and creatively about how to extend existing concepts and principles and perhaps, more radically, devise the new ones we and AI need.

Engineering a 'Future of Work': The Politics and Ethics of Robotics and AI Research
Yunus Dogan Telliel, *Worcester Polytechnic Institute, United States*

Current public debates around robots and other autonomous systems succumb to either techno-pessimism and techno-optimism. While proponents of the first believe that the adoption of these systems will eventually move large populations out of workforce and rob them of their ability to support themselves and their sense of identity and moral worth, proponents of the latter consider the very same trend as the liberation of humans from the necessity of work. Yet, some North American engineering researchers still operate with the assumption that in the foreseeable future workplaces will remain human-centric, and human control, input, and guidance will shape the success of human-robot collaboration. Although engineers play a crucial role in the design, development, and delivery of work-related robotic technologies, the 'future of work' debate tends to overlook their ethical commitments and future imaginations. Drawing on an ongoing research project with engineering researchers, this paper discusses how engineers articulate (and—sometimes—fail to articulate) the future implications of their work for themselves, their colleagues, and the broader public. My discussion will focus on two questions: 1) in what ways can and do engineers move beyond a narrowly-framed calculus of jobs created or made obsolete, and think about the quality, security, meaning of work in new technology ecosystems?, and 2) to what extent do they align their research and design practices with an inclusive and equitable vision of the future workplace?

War or Peace Between Humanity and Artificial Intelligence
https://journals.library.iit.edu/index.php/CEPE2023/article/view/268
Wolfhart Totschnig, *Universidad Diego Portales, Chile*

The thinkers who have reflected on the potential risks of a future artificial general intelligence (AGI) have focused on the possibility that the AGI could incidentally destroy our world, and consequently, us, because it misinterprets the goal that we have given to it (Yudkowsky, Bostrom, Omohundro, Yampolskiy, Tegmark, Russell). They have neglected the possibility that the AGI could come to see us as a threat to its existence and, therefore, deliberately try to eliminate us. The aim of the present paper is to show that this neglect is mistaken. I will describe a possible situation where an AGI and humanity find themselves vulnerable vis-à-vis each other, which could lead to an all-out war. I will then argue that, in view of this possibility, the approach of the said thinkers, which is to search for ways to keep an AGI under control, is potentially counterproductive because it might, in the end, bring about the existential catastrophe that it is meant to prevent.

12:15–1:30 p.m. **Lunch**

1:30–3 p.m. **Plenary Session**
Hermann Hall Ballroom

Machine-Readable Humanity: What's Wrong With That?
Helen Nissenbaum*, Cornell University, United States*

Helen Nissenbaum is a professor of Information Science and founding director of the Digital Life Initiative at Cornell Tech, New York City. Her work focuses on ethical, and political implications of digital technologies on issues such as privacy, bias in digital systems, trust online, ethics in design, and accountability in computational and algorithmic systems. Prof. Nissenbaum's publications, which include the books, Obfuscation: A User's Guide for Privacy and Protest, with Finn Brunton (MIT Press, 2015), Values at Play in Digital Games, with Mary Flanagan (MIT Press, 2014), and

Privacy in Context: Technology, Policy, and the Integrity of Social Life (Stanford, 2010), have been translated into seven languages, including Polish, Chinese, and Portuguese. Grants from the NSF, Air Force Office of Scientific Research, the U.S. Department of Health and Human Services Office of the National Coordinator, McArthur Foundation, DARPA, and NSA have supported her research. Recipient of the 2014 Barwise Prize of the American Philosophical Association and the IACAP Covey Award for computing, ethics, and philosophy, Prof. Nissenbaum has contributed to privacy-enhancing free software, TrackMeNot (protecting against profiling based on Web search) and AdNauseam (protecting against profiling based on ad clicks). She holds a Ph.D. in philosophy from Stanford University and a B.A. (Hons) in Philosophy and Mathematics from the University of the Witwatersrand, South Africa.

Abstract:

As our lives become increasingly managed by, and mediated through complex digital systems, it's taken for granted that we need to be "machine readable."However, in defending tools of obfuscation such as *AdNauseam*, which resist machine readability by muddying the data pool, Howe and Nissenbaum argue that there is no obligation to facilitate the classification of humans in service of arcane, automated systems of programmatic advertising. Being machine readable, according to this picture, means being dehumanized as cogs in automated systems that sustain a web of surveillance and control.

My talk, based on work-in-progress with Solon Barocas and Margot Hanley, addresses whether machine readability, itself, is a moral problem, or is a problem for reasons particular to the case at hand. As a first step toward answering this question, we develop a conception of machine readability that is both coherent and generalizable and draw on this conception to step through progressive questioning of diverse systems that "read" humans in different ways. This progressive questioning yields a set of moral and nonmoral dimensions of these systems that influence how we evaluate them in moral terms, and, in turn, how we may estimate the degree of our obligation to be legible to them.

3–3:15 p.m. **Break**

3:15–4:45 p.m. **Session 2**

Room 1 - Regulation of AI

Hermann Hall Ballroom
Chair: Ray Trygstad

Where Law and Ethics Meet: A Systematic Review of Ethics Guidelines and Proposed Legal Frameworks on AI

https://journals.library.iit.edu/index.php/CEPE2023/article/view/282

Désirée Martin and Michael W. Schmidt,
Karlsruhe Institute of Technology, Germany

From an ethical perspective, there have been efforts to systematize the emerging number of AI ethics guidelines. Nevertheless, there is no comparison of relevant ethics guidelines with current regulatory proposals. Our paper aims to fill this research gap.

Methodologically, we focus on relevant keywords that are inconsistently classified across the texts. Aiming for unification, we categorize the relevant keywords systematically as values or principles and extract their relations and hierarchies.

Besides providing a systematic and relational list of moral values and principles, which are widely shared in the moral and legal realm, we also provide explications of these values and principles based on our synoptic findings. This serves to foster a better understanding of these values and principles, which is essential for assessing the acceptability of AI technology and its application.

A further finding is that in the current legal proposals, not every shared ethical principle or value is already included. This can be useful for the further development of the legal system by illustrating important ethical aspects in AI that are not (yet) transferred to legal guidelines.

Governance Conflicts and Public Court Records
https://journals.library.iit.edu/index.php/CEPE2023/article/view/230
Kyra Milan Abrams and Madelyn Rose Sanfilippo,
University of Illinois at Urbana-Champaign, United States

Datafication of society has heavily influenced the way in which we use technology and how technology is designed. Social informatics research argues that the way in which technology is used by different groups is not neutral. With the increased usage of technology, the data users provide has also been a topic of research. Furthering that argument, data governance not only differs, it also is influenced by those who have the power to control it. With the differences in how data is governed differing across data, what does that mean for data that is used to train models? How do the implications of data governance shape how what is trained? This paper seeks to evaluHuman cate that relationship through multi-method content analysis of governance documents regarding data and access to public court records in Illinois and California. It seeks to fill the gap in research surrounding ethical impacts of data governance and how those impacts can have larger and possibly negative implications.

Framing Effects in the Operationalization of Differential Privacy Systems as Code-Driven Law
https://journals.library.iit.edu/index.php/CEPE2023/article/view/264
Jeremy Seeman, *Pennsylvania State University, United States*

As privacy regulation evolves to strengthen protections for data subjects, data processors seek clarity on how to comply with these new regulations. Differential privacy (DP), a mathematical framework for assessing personal data privacy risks, has been proposed to help delimit the bounds of personal data and help processors comply with privacy law. But like all technologies, DP exhibits framing effects, in that the way DP defines and manages privacy harms legitimizes certain sociotechnical interventions and delegitimizes others. This paper investigates DP's framing effects and their political implications for using DP as a code-driven instrument of privacy policy.

We describe how DP's framing strengthens some substantive privacy protections while eschewing other sociological dimensions of privacy, potentially modulating data subject rights and the power of auditing organizations. In doing so, we propose how to delineate where new interventions are needed for regulating DP systems while still harnessing the power of DP's privacy guarantees to protect data subjects from potential harms.

Room 2 - AI Agency

Hermann Hall Expo

Chair: Susan Dwyer

Does it (Morally) mMatter Whether the AI Machine is Conscious?

https://journals.library.iit.edu/index.php/CEPE2023/article/view/239

Kamil Cekiera, *University of Wroclaw, Poland*

A rapid pace of the AI technologies development is on par with a growing interest in issues related to the AI's functions, status, and influence on the society. In the face of it, philosophers needed to take them into account as well. Recently philosophers began to pay attention to the role of consciousness in their theorizing about the AI ethics. According to the view defended by David Chalmers (2022), if the AI machines could be conscious and that would grant them moral status comparable to that of human beings, that changes the way we should think not just about the moral status of artificial being, but about the concept of human itself. As I am going to show,
if Chalmers is right, then the moral standing of robots is not just similar to that of humans, but instead robots should be considered humans. Thus, in such case we would need to engineer the concept of human.

In my talk I am going to show how Chalmers' argumentation inevitably leads to that conclusion, elucidate what functions we want the concept of human to fulfill, and argue that Chalmers' argumentation is flawed as it is not clear that conscious machines are possible.

Do We Have Procreative Obligations to AI Superbeneficiaries?

https://journals.library.iit.edu/index.php/CEPE2023/article/view/290

Sherri Conklin, *University of California Santa Barbara, United States*

This paper considers our obligations to AI superbeneficiaries – entities with inherently valuable interests that exceed those of humans in terms of quality and/or quantity. The issue at stake is that AI superbeneficiaries potentially threaten the well-being of humans. If we do bring AI superbeneficiaries into existence, then we will have moral obligations to promote the interest of these AI at the expense of human interests. If we have these obligations, do we have further moral obligations to bring such entities into existence? By applying an anti-natalist argument, this paper argues that we have moral obligations against bringing AI superbeneficiaries into existence because of the existential risk they pose to their own survival.

Can AI Determine its Own Future?

Aybike Tunc, *Ankara Hacı Bayram Veli University, Turkey*

In the previous days, one of Google's engineers, Blake Lemoine, published an interview with an AI system called LaMDA claiming that it is sentient and in fact, a person.

LaMDA may be the first AI system that claims personhood, but it will certainly not be the last one. The main issue here is how the law will respond to that claim. Of course, LaMDA was referring to moral personhood but being a person has legal consequences too.

LaMDA's claim was an example of self-determination. Self-determination, or autonomy, is a legal right based on being independent. In its simplest definition, being independent means not being subject to the control of another. In other words, it means having free-choice.

The AI systems that exist today are under the supervision of a person. However, LaMDA and AlphaGo examples showed us that it can be possible for, even with a single move, an under-control AI to have free-choice. For this reason, the law should prepare itself for future AI technologies that may act independently according to its free-choice.

Day 2 - May 17

9–10:30 a.m.

Session 3

Room 1 - Deepfakes and Hate Speech

Hermann Hall Ballroom
Chair: Ray Trygstad

Foundation Models, Forgeability, and Evidence in Politics
https://journals.library.iit.edu/index.php/CEPE2023/article/view/280
Megan Hyska, *Northwestern University, United States*

Deep learning models that can create novel, high-quality audio-visual samples--- deepfakes--- are already with us, and will become more accessible to non-specialist users in the coming years. This paper draws out a political consequence of what we must anticipate will be the resulting ubiquity of fake videos: it will make it harder to engage in showing, as opposed to telling, in long distance communication, and so will present a challenge to the process of online political organizing. Showing is the sort of communication that presents its audience with evidence independent of the speaker's inferred intentions, and it is therefore a uniquely powerful sort of communication in the context of persuading those who don't already trust the speaker. Departing from recent work by Don Fallis, I suggest that the ubiquity of deepfakes will modify political information flow in kind rather than degree. And departing from work by Regina Rini, I emphasize the epistemic challenges of deep fakes in the context of extra-institutional politics.

Deepfakes and Dishonesty
https://journals.library.iit.edu/index.php/CEPE2023/article/view/242
Tobias Flattery and Christian Miller, *Wake Forest University, United States*

Deepfakes raise various reasons for concern. However, there has been almost no sustained philosophical analysis of deepfakes from the perspective of the philosophy of honesty and dishonesty. Obviously deepfakes are potentially deceptive. But under what conditions does using deepfakes fail to be honest? Which agents involved are dishonest, and in what ways? To understand better the morality of deepfakes, these questions need answering. Our first goal, therefore, is to offer an analysis of paradigmatic deepfakes in light of the philosophy of honesty. There are, of course, reasons to think that deepfakes could supply or support moral goods. Even so, it doesn't follow that these uses of deepfakes are honest. Our second goal, therefore, is to apply our analysis of deepfakes and honesty to the sorts of deepfakes hoped to be morally acceptable. We conclude that in many of these cases the use of deepfakes will be dishonest in some respects.

Improving AI-Mediated Hate Speech Detection: A Genuine Ethical Dilemma
https://journals.library.iit.edu/index.php/CEPE2023/article/view/289
Maren Behrensen, *University of Twente, Netherlands*

AI-mediated hate speech detection is indispensable for contemporary communication platforms. But it has known deficiencies in terms of bias and context-awareness. I argue that improving on these known deficiencies leads into a genuine ethical dilemma: It will increase the epistemic and social utility of these platforms, while also helping bad faith political and corporate actors to suppress unwelcome speech more swiftly and efficiently.

Room 2 - Moral Frameworks

Hermann Hall Expo

Chair: Katleen Gabriels

AI Ethics: A Perspective from American Pragmatism

https://journals.library.iit.edu/index.php/CEPE2023/article/view/262 Andréane Sabourin Laflamme and Frédérick Bruneault,
André-Laurendeau College, Canada

Throughout the history of moral philosophy, the theoretical postures have been privileged. Modern ethics is no exception and is indeed characterized by the predominance of voluntarist and universalist frameworks, which are primarily concerned with the actions of the moral agent, with no real regard for the conditions of possibility necessary for the effective realization of moral actions. Recent developments in applied ethics have shown that an integral application of classical ethical frameworks does not adequately address the new moral dilemmas emerging from our different spheres of activity. Artificial intelligence (AI) ethics once again demonstrates the inadequacy of traditional ethical frameworks to deal with the many ethical issues related to the pervasiveness of AI systems. Indeed, the dominant theories in ethics fail to take account of the shared responsibility that characterizes the moral obligations we have towards AI systems. The particularity of pragmatist ethics is that it aims at a practical intervention without however renouncing the conceptual clarifications necessary for such an intervention. We will demonstrate how the characteristics of pragmatist ethics avoids certain pitfalls in AI ethics and provides a conceptual framework particularly well suited to address the ethical issues related the increasing use of AI systems in our societies.

What is AI Ethics? Why Codes of Conduct and Normative Claims Need Ethical Reflection

https://journals.library.iit.edu/index.php/CEPE2023/article/view/227
Suzana Alpsancar, *Paderborn University, Germany*

In this paper, I highlight three common meanings of ethics that can be found in the field of AI research but are not consistently distinguished: (1) Ethics as a principle of self-regulation, (2) ethics as an attribute, and (3) ethics as a process of reflection and deliberation. I argue that the idea of self-regulation that underlies the countless recently published guidelines for AI development and business rests on the presumption that those who commit themselves to self-regulation must be equal to another in regard to the subject of self-regulation. Moreover, ethical guidelines need ethical deliberation, even for those who authentically commit themselves to these principles. The same holds true for using 'ethical' as a qualifying attribute. Accordingly, the AI community should think more about what frameworks can be conducive to cultivating ethical reflection and deliberation.

Humanity Compatible: Aligning Autonomous AI with Kantian Respect for Humanity

https://journals.library.iit.edu/index.php/CEPE2023/article/view/292
Ava Thomas Wright, *California Polytechnic State University, San Luis Obispo*

In this paper, I will argue that autonomous AI agents designed along Russellian lines should be programmed to determine our objectives by modeling our agency substantively as Kantian autonomy, rather than as the satisfaction of preferences. The objectives that the AI infers from our behavior should not be understood in terms of efforts to satisfy preferences but instead in terms of efforts

to act autonomously in the Kantian sense of that term. Only by modeling us as autonomous agents will the AI be able to learn and help us to achieve our objectives. Any other model of our agency would fail to treat us as ends and so fail to respect our humanity. I thus argue for humanity compatible AI.

10:30–10:45 a.m. **Break**

10:45 a.m.–12:15 p.m. **Plenary**
Hermann Hall Ballroom

Metaverse Ethics: Foundations and Key Issues
Philip Brey, *University of Twente*

Philip Brey (PhD, University of California, San Diego, 1995) is professor of philosophy and ethics of technology at the Department of Philosophy, University of Twente, the Netherlands. He is currently also programme leader of the ESDiT (Ethics of Socially Disruptive Technologies), a ten-year research programme with a budget of € 27 million and the involvement of seven universities and over sixty researchers (www.esdit.nl). Esdit runs from 2020 to 2029. He is a former president of the International Society for Ethics and Information Technology (INSEIT), and of the Society for Philosophy and Technology (SPT). Hi is also former scientific director of the 4TU. Centre for Ethics and Technology 2013-2017. He is on the editorial board of twelve leading journals and book series in his field, including Ethics and Information Technology, Nanoethics, Philosophy and Technology, Techné, Studies in Ethics, Law and Technology and Theoria.

Abstract:

The metaverse is a network of immersive, persistent, interoperable virtual worlds that serve as a platform for social interaction, entertainment, commerce, work, education, healthcare, industrial production and other functions. In this talk, the case will be made that at some point in the not too distant future, the internet will in large part consist of a metaverse-like environment, and it will be examined what new ethical issues will emerge in this new constellation. It will be argued that a major ethical shift will be needed in our thinking about the internet, from an ethics of information, communication and media to an ethics of embodied interaction with world-like simulated objects and environments. This will require a new ethics of embodied virtual interaction, an ethics of virtual actions and events, and an ethics of design and governance of virtual worlds. The case will be made that metaverse ethics, thus conceived, differs substantially from current digital ethics and from an ethics of physical interactions and events, and that its development will require a rethinking of current moral concepts and theories. Security in the metaverse is more than security of data and computer resources; it also pertains to security of (avatar-mediated) persons. Privacy in the metaverse is more than information privacy: it is also bodily and spatial privacy. Property rights need to be rethought as well, in relation to virtual property, NFTs and other metaverse assets. It will be concluded that while the metaverse may bring substantial benefits, it will also raise a series of new ethical issues and challenges, and will engender ethical risks that far exceed those of the current internet. A fundamental consideration of ethical issues is therefore needed right from the onset of its development.

12:15–1:30 p.m. **Lunch**

1:30–3:00 **Session 4**

Room 1 - Datafication and the Digital Self

Hermann Hall Ballroom
Chair: Fran Grodzinsky

Understanding Freedom in the Age of the Machines: What does it Mean to Be Digitally Free?

https://journals.library.iit.edu/index.php/CEPE2023/article/view/276
Migle Laukyte, *Pompeu Fabra University, Spain*

The XXI century is a century of disruptive technologies: How these technologies will affect our freedoms is not clear. However, we exercise our freedoms not only when we are inside the social network, use apps or navigate the Internet: we exercise freedom also when we choose not to use any of these technologies. This paper is about such an understanding of freedom. The question is whether and to what extent the freedom not to use technologies is real. The freedom from technologies has been articulated in a variety of ways, among which is the language of rights. I look at some of these rights—coming from personal data protection, labor law and administrative law domains—and argue that they represent a particular shape of freedom from technologies that we still are willing to guarantee to humans.

The Digital Alienation from The Self: An Epistemic Argument

Damian Fisher and Syed Abumusab,
University of Kansas, United States

This paper argues that digital technologies present new kinds of alienation which require the introduction of the concept *digital alienation*. By "digital alienation," we mean the unique kind of alienation caused by digital technologies, especially the overuse of digital technologies. We provide a robust framework for digital alienation, by defining "digital alienation," introducing three novel kinds of digital alienation, and then by explaining the reciprocal and reinforcing relationship between these three novel kinds of digital alienation. From this, we argue these three kinds of digital alienation cause a feedback loop and this feedback loop compounds the negative effects of digital alienation. The upshot of our argument is three-fold: (i) we distinguish the Hegelian-Marxist "alienation" from "digital alienation," (ii) we provide a causal account of "digital alienation," and (iii) we provide a concept that captures feelings individuals' are already having but are unable to conceptualize.

Data After Death: Remembrance and Resurrection

https://journals.library.iit.edu/index.php/CEPE2023/article/view/278
Alexis Elder, *University of Minnesota, United States*

How should we engage with people's data after death? Recent work in information ethics has urged us to consider data about the dead as analogous to physical remains, and subject to relevant norms about respectful handling. At the same time, information ethicists have emphasized the social nature of data: data about me is also data about my friends, loved ones, relationships and community, making it difficult to assign ownership to individual bits of data. Privacy theorists have urged us to move beyond a simplistic public/private model of information flow, emphasizing the importance of context and relationality in thinking about norms around privacy. I draw on Zhuangzi's philosophical

accounts of interconnectedness to develop the analogy to physical remains in a way that can help us move beyond individualist accounts of data stewardship for the dead, and reflect on the roles data can play in remembrance.

Room 2 - AI in Healthcare

Hermann Hall Expo

Chair: Valentina Beretta

Psychotherapist Bots: Transference and Countertransference Issues

https://journals.library.iit.edu/index.php/CEPE2023/article/view/22

Saeedeh Babaii, *University of Tuebingen, Germany*

In recent years, the field of ethics of artificial intelligence has been majorly discussing the concept of trust and what it means when we use notions like trust and trustworthiness for AI systems. Multiple theoretical frameworks, methodologies, and conceptions have been developed and recruited to articulate the concept of trust in AI ethics literature e.g., XAI. In this paper, I plan to enrich the concept of trustworthy AI with the well-established literature on care ethics and to investigate the role of decision-subject's emotions in shaping a trustworthy network of actors. To make the proposed thoughts and ideas more applicable, I will focus on a concrete case study namely AI-assisting clinical treatment and address questions like how the network of trust relations and the process of clinical decision-making change in such settings. I will explore how my proposed conception of trust relations can contribute to a more trustworthy AI-consulted treatment planning.

Artificial Intelligence in Healthcare: An Analysis of Training Needs in Europe

Valentina Beretta, *University of Pavia, Italy*
Maria Chiara Demartini, *University of Pavia, Italy*
Hatim Abdulhussein, *Health Education England*
Marco Fisichella, *Leibniz University, Hannover, Germany*
Franziska Schoger, *Leibniz University, Hannover, Germany*
Dennis Vetter, *Goethe University, Frankfurt, Germany*
Blaz Zupan, *University of Ljubljana, Slovenia*
Ajda Pretnar, *University of Ljubljana, Slovenia*

It is imperative for public and private investments in AI, to prepare for socio-economic changes, and ensure an appropriate ethical and legal framework for its use. The aim of this paper is to draw up the needs assessment for a curriculum proposal on the topics of AI, healthcare management and ethics, including all the necessary agreements, as well as the identification of the respective profiles and competencies of health managers and the strategies that may be agreed upon for the production of curricula. In particular, this paper aims at mapping the vision and needs of all the possible different actors and stakeholders that could be interested. This study adopts a mixed method approach with three different data collection methods. Online survey, virtual interviews and stakeholder(s) workshops were undertaken. The analysis is conducted in Europe. According to the results, five main areas emerged as particularly important when designing educational proposals for future healthcare workforce. These are inclusive of the human factor and attention to the patient, the implications derived by workforce changes, the limitations associated with data collection and analysis, the ethical and legal considerations and the need of data translators. This study raises awareness on the topic at national and regional level by providing further evidence on the identification, analysis and systematization of the different stakeholders and by addressing potential workforce needs with specific regards to the healthcare sector at different levels of analysis.

In line with empirical research, this study is not without limitations, which open avenues for future research.

Epistemic Injustice and Algorithmic Epistemic Injustice in Healthcare
https://journals.library.iit.edu/index.php/CEPE2023/article/view/238
Jeffrey Byrnes and Andrew Spear,
Grand Valley State University, United States

We argue that the introduction of algorithmic support systems into medical decision-making, while holding out much promise, also exacerbates ethical concerns deriving from existing knowledge- and power-asymmetries between healthcare providers on the one hand and patients on the other. In several areas, issues with which medicine is already struggling threaten to become more ethically fraught as algorithmic systems enter the picture. Worse still, the very authority accorded to such systems might serve to cover over or render invisible this fraughtness. Recent literature argues that epistemic injustices, harms to individuals in their capacity as knowers, are particularly likely to occur in the healthcare context due to unwarranted but common prejudices concerning the insights and self-understanding of the ill. Testimonial injustice, for example, where a patient's report or viewpoint concerning their own condition is not taken seriously due stereotypes of ill persons as uninformed or cognitively compromised. We argue that, at least in cases where patients already suffer an unwarranted credibility deficit and so injustice of this sort, the deployment of algorithmic systems is likely to reinforce and amplify the credibility deficit and so injustice due in part to the elevated objectivity and credibility ("automation bias") often ascribed to such systems.

3–3:15 p.m. **Break**

3:15–4:45 p.m. **Session 5**

Room 1 - Virtual Reality and the Digital Space
Hermann Hall Ballroom
Chair: Kelly Laas

Theoretical Underpinnings of Virtual Reality: From Second Life to Meta
https://journals.library.iit.edu/index.php/CEPE2023/article/view/279
Katleen Gabriels, *Maastricht University, Netherlands*

Since Facebook's transition and rebranding to 'Meta' in October 2021, there is a renewed academic and societal interest in the notions of 'metaverse,' 'virtual reality' (VR), and 'virtuality' (see e.g., Novak, 2022; Gent, 2022). This renewed interest reminds of the debates around the three-dimensional social virtual worlds Second Life in 2007. This paper has a two-fold conceptual aim. First, it presents a critical synthesis of how late-twentieth and twenty-first century philosophers and media theorists have conceptualized virtuality and its relation to reality, in the context of VR. The analysis carefully distinguishes seven theories. The second part focuses on a comparison (similarities and dissimilarities) between Second Life and Meta. The starting points are four conceptualisations of virtuality: an ontological, a phenomenological (in terms of subjective embodied experience), a cultural, and a technological conceptualization (e.g., VR; augmented reality). Ultimately, both aims and parts seek to contribute to a better and more nuanced understanding of the theoretical underpinnings of the current academic and societal discussions about Meta.

XR Embodiment and the Changing Nature of Sexual Harassment
Erick Ramirez, Shelby Jennett, Raghav Gupta, *Santa Clara University, United States*
Jocelyn Tan*, Sisu VR*

New technologies introduce novel or enhanced forms of communication and can transform the nature of social interaction in often unpredictable ways. In this paper we assess the impact of extended reality technologies as they relate to sexual forms of harassment. We begin with a history of sexual harassment and then offer an account of extended reality technologies focusing specifically on the psychological and hardware elements of XR. Although virtual spaces of different forms exist (private, semi-private, and public spaces), we focus on public social spaces in order to explain how the concept of sexual harassment must be adjusted. We then offer a typology of sexual harassment for the metaverse focusing on three distinct forms of sexual harassment: 1) invariant forms of harassment (forms of harassment that function identically in physical, online, and metaverse contexts); 2) extended forms of harassment (harassment that takes on new forms in the context of the metaverse); and 3) metaverse specific forms of harassment (harassment that arises only due to the unique psychological or hardware features of metaverses).We argue that existing frameworks will not helpfully address metaverse-specific harassment.

An Investigation in the (In)Visibility of Shadowbanning
https://journals.library.iit.edu/index.php/CEPE2023/article/view/259
Amanda Pinto, *Marquette University, United States*

Social media, such as Instagram, rely on curation algorithms for user's feeds and content moderation algorithms to hide inappropriate, violent content. These algorithms work to create both visibility and invisibility of bodies based on trained policies of acceptability and appropriateness. Within the liminal area of what is appropriate, where posts or users do not clearly violate community guidelines, is the technique of shadowbanning. Shadowbanning as a practice by Instagram is only known by those who experience it, as Instagram continues to deny its existence within the algorithm. Yet the presence of users whose engagement becomes limited, the reach almost nonexistent, and discoverability low, all seen through Instagram's own analytics, suggest a technique of invisibility that presents an illusion of visibility within the algorithm. To explore the technique of shadowbanning, I will focus on the community of recreational pole dancers and their proximity to what is deemed as "inappropriate" nudity.

Day 3 - May 18

9–10:30 a.m. **Session 6**

Room 1 - Interacting with AI in Social/Emotional Contexts
Hermann Hall Ballroom
Chair: Alexis Elder

Beyond Turing: Ethical Effects of Large Language Models
https://journals.library.iit.edu/index.php/CEPE2023/article/view/244
Alexei Grinbaum and Laurynas Adomaitis, *CEA-Saclay, France*

Manufacturers often present large language models (LLMs) as "personal assistants" or "virtual friends". Despite being almost always aware that they are talking to a machine, users perform spontaneous projections and, over

time, begin to speak with machines as they do with humans. We argue that the indistinguishability of linguistic performance between humans and machines is no longer a key issue. It is not required for psychological, emotional, moral, or social effects to take place. Instead, the concern is whether the effects of machine language on human users are equivalent to the ones experienced in a human-to-human interaction, even when machine outputs are marked as such. To study this “beyond Turing” regime, we use two case studies, then offer a classification of effects that persist even in the case of full cognitive awareness. We conclude that these lasting effects of language-generating machines imply a duty of reflective design and care.

Sex-bots and Touch: What Does it All Mean for our (Human) Identity?
https://journals.library.iit.edu/index.php/CEPE2023/article/view/233
Iva Apostolova, *Dominican University College, Canada*

In this paper I will be exploring the significance of touch in relation to human/personal identity. As a point of illustration, I will be using an unusual angle, namely sex-bots and their place in human sexuality. Sex-bots present a unique challenge since their purpose is mainly, if not exclusively, to engage in tactile interaction of sexual nature, broadly construed. In this sense, I will use them as a theoretical decoy to explore connections between the sense of self and the faculty of touch. My claim is that the formation of human-type consciousness requires the faculty of touch, which in turn, is central for the development of feelings such as compassion and empathy, both of them at the heart of any (human) relationship. The faculty of touch is at the foundation of the multisensory integration process involved in human perceptivity. In this sense, I will argue in favor of a bottom-up construction of the (human) self.

Can Large Language Models as Chatbots be Social Agents?
https://journals.library.iit.edu/index.php/CEPE2023/article/view/232
Syed Abumusab, *University of Kansas, United States*

In this article, I discuss the concept of social agency in AI systems, specifically chatbots. I argue that, like agency, social agency is not a matter of meeting some threshold of human-like capacities. Instead, it is a matter of degree, and chatbots can be social agents to some degree. I propose a theoretical framework for chatbots’ social agential status, which also specifies the conditions a chatbot should satisfy to be considered a proper implementation as a social agent. I emphasize the importance of staying sensitive to existing social theories while leaving the possibility of tweaking them to account for new social facts instantiated by human-AI system relationships.

To defend chatbot social agency, I deploy the idea of levels of abstraction (LoA) made popular by Luciano Floridi, which allows one to focus on a particular aspect of the domain of inquiry. I argue that chatbots can be seen as social agents when viewed at a particular LoA, namely conversational, linguistic, or social LoA. At this LoA, they can exhibit a dimension or degree of sociality and participate in social activities appropriate for chatbots. There may be a mismatch between existing social theories and chatbots, but I argue that a pluralistic and updated sociality framework is better equipped to account for this phenomenon.

Overall, the article highlights the importance of recognizing the social-agential status of AI systems and the need for a nuanced understanding of social agency in these systems.

Room 2 - Decision Support

Hermann Hall Expo

Chair: Kelly Laas

When can a Decision Support System Nudge?

https://journals.library.iit.edu/index.php/CEPE2023/article/view/240

Francesco Pedrazzoli, Fabio Aurelio D'Asaro and Massimiliano Badino
University of Verona

The literature generally acknowledges that we should refrain from delegating important decisions to machines, especially in high-risk fields, as recognized, e.g., in the Artificial Intelligence Act. These documents rule out the use of Decision Support Systems (DSSs) based on Artificial Intelligence (AI) for automatic decision-making. In other words, humans must always participate in the decision-making pipeline (this is sometimes called the human-in-the-loop model). However, this does not help when we are put in the very practical situation where we must decide what portion of our decisional power we are willing to share with a DSS. The picture gets muddier when we consider that DSSs can operate some form of nudging on the user. In this position paper, we analyze the concept of support within the field of Human-AI Interaction and its relation to nudging. We propose that the nudging component should be evaluated when building ethical DSSs. To this aim, we argue that characteristics of a DSS, such as its bias and transparency, should be considered to evaluate whether nudging is legitimate. We provide a coarse-grained taxonomy of DSSs and nudging to facilitate such an evaluation.

Rebalancing the Digital Convenience Equation through Narrative Imagination

https://journals.library.iit.edu/index.php/CEPE2023/article/view/271

Fernando Nascimento and Anya Workman, *Bowdoin College, United States*

Digital technologies offer increasing conveniences for our daily lives. Facilitated by recommender systems, we buy things without leaving the house, and listen to music in continuous playlists that entertain us without intervention. These AI-enable conveniences captivate our society, making us increasingly enamored with digital technologies. In these conveniences, there are at least two fundamental underlying characteristics that eclipse ethical considerations: immediacy and egocentrism. However, in recent years such conveniences have begun to take their toll, with disruptions in democratic systems, amplified social discrimination, alarming increases in teenage suicides, and the growth of digital-empowered economic and social inequality. Thus, new questions are raised concerning the values, responsibilities, and freedoms associated with AI- based systems.

The imaginative power of narratives has long mediated the ethical reflection on the long-term and social consequences, counterbalancing our tendency towards the constant and immediate realization of what is pleasurable. Based on hermeneutics methodology and Paul Ricoeur's narrative theory, we argue that systematic exposure to fictional and historical narratives can break the monopoly of immediate and egocentric convenience, creating a space for reflection that expands the decision-making process of using and adopting new technologies to consider otherness and long-term implications.

10:30–10:45 a.m. **Break**

10:45–12:15 a.m. **Session 7**

Room 1 - Autonomous Technology

Hermann Hall Ballroom
Chair: Ray Trygstad

People's Perception and Expectation of Moral Settings in Autonomous Vehicles: An Australian Case

https://journals.library.iit.edu/index.php/CEPE2023/article/view/284

Amir Rafiee, Hugh Breakey, Yong Wu and Abdul Sattar, *Griffith University, Australia*

While Autonomous Vehicles (AVs) can handle the majority of driving situations with relative ease, it is indeed challenging to design a system whose safety performance will fit every situation. Technology errors, misaligned sensors, malicious actors and bad weather can all contribute to imminent collisions. If we assume that the wide-spread use and adoption of AVs is a necessary condition of the many societal benefits that these vehicles have promised to offer, then it is quite clear that any reasonable ethics policy should also consider the various user expectations with which they interact, and the larger societies in which they are implemented. In this paper we aim to evaluate Australian's perception and expectation on personal AVs relating to various ethical settings. We do this using a survey questionnaire, where the participants are shown 6 dilemma situations involving an AV, and are asked to decide which outcome is the most acceptable to them. We have designed the survey questions with consideration for previous research and have excluded any selection criteria which we believed were biased or redundant in nature. We enhanced our questionnaire by informing participants about the legal implications of each crash scenario. We also provided participants with a randomised choice which we named an Objective Decision System (ODS). If selected, the AV would consider all possible outcomes for a given crash scenario and choose one at random. The randomized decision is non-weighted, which means that all possible outcomes are treated equally. We will use the survey analysis, to list and prioritize Australian's preferences on personal AVs when dealing with an ethical dilemma, that can help manufacturers in programming and governments in developing AV policies. Finally, we make some recommendations for further researchers as we believe such questionnaires can help arouse people's curiosity in the various ways that an AV could be programmed to deal with a dilemma situation and would encourage AV adoption.

Toward Substantive Models of Rational Agency in the Design of Autonomous AI
https://journals.library.iit.edu/index.php/CEPE2023/article/view/286
Ava Thomas Wright and Jacob Sparks, *California Polytechnic State University, San Luis Obispo*

Artificially intelligent autonomous agents are “autonomous” in the sense that they are programmed to learn for themselves how to act across a wide range of situations. What they will do in any given situation, therefore, cannot be completely foreseen or predicted in advance. AI autonomy gives rise to the problem of value alignment, How can we make sure that AI agents, acting autonomously, will behave in ways that align with moral values? In this paper, we will argue that this problem cannot be solved so long as AI rational agency is conceived strictly instrumentally. A more substantive conception of rational agency is needed, one in which autonomous machine agents reason not only about how to efficiently achieve their ends, but also about what those ends should be.

Room 2 - AI in Healthcare: Data Management and Cybersecurity

Hermann Hall Expo
Chair: Elisabeth Hildt

A Labor History of Health Records: On Medical Scribes and the Ethics of Automation
https://journals.library.iit.edu/index.php/CEPE2023/article/view/265
Sara Simon, *Illinois Institute of Technology*

This paper explores the human labor demands that underpin the utility of patient health records. I examine where these labor demands originated historically, and I consider how they might evolve, given the recent rise of artificial intelligence (AI) being developed to automate the collection and categorization of patient health information. Using a sociotechnical framework, the paper identifies a complicated paradox: the labor of medical scribes has become crucial for the benefits of electronic health records (EHR) to be realized; simultaneously, scribe work has been regarded in medical literature as inconspicuous and transitory, a stopgap measure wholly replaceable by a more efficient solution. The paper thus critically interrogates the premise that automation can replicate and replace scribe labor, examining the ethics of moving toward a fuller reliance on AI.

Automation, Trust, Responsibility in Algorithmic Warfare
https://journals.library.iit.edu/index.php/CEPE2023/article/view/248
Stefka Hristova, *Michigan Technological University, United States*

In a 2006 editorial for National Defense, Stew Magnuson made an apt observation: "The robot army is coming." Algorithmic war has been envisioned as war fought by algorithmic technology under the guise of protecting human life and in response to a potential enemy robot army. As David Humbling has reported, in preparation for this new war, "[o]ne U.S. Navy project envisages having to counter up to a million drones at once" (Hambling 2021). The algorithmic technology developed is indeed one that envisions both attacks and counterattacks as air combat. The military's robot army increasingly consists of autonomous technology deployed on jets and drones. In 2020, the "U.S. Air Force let an artificial intelligence take over the navigation and sensor systems of a Lockheed U-2 spy jet during a training flight [marking] the first known time an AI has to been used to control a US military aircraft" (The Airforce 2020). Here, onboard the U-2 "Dragon Lady" spy plane, the "human Air Force officer" was partnered with "ARTUμ algorithm" which is now responsible for real-world sensor monitoring and navigation and yet is modeled after a gaming system (Browne 2020). While these seem like small, incremental steps toward algorithmic war, they point to an ambitious goal where in "10 to 15 years max, you are going to see the widespread, ubiquitous use of robots throughout most militaries in the world"(2020). This idea of robot-driven warfare has been met with skepticism as it raises significant moral and ethical issues about trust and responsibility.

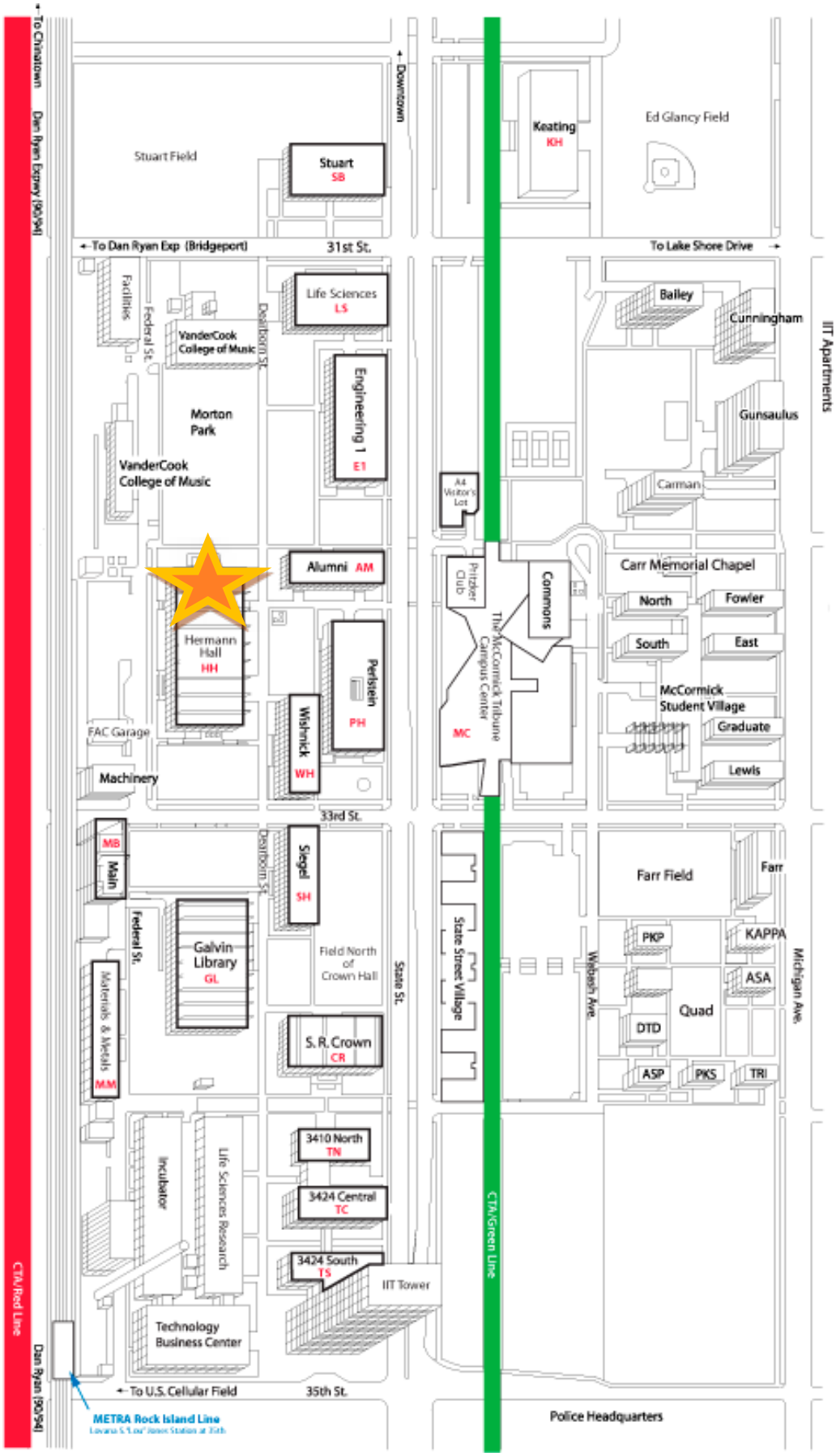
Trust War systems are increasingly seen as entirely unmanned and thus autonomous. The processes of automation of war require the articulation of three major interrelated processes as they relate to trust. As Paul Scharre has aptly written, "Activating an autonomous system is an act of trust" (2018, 149). First, the process of building trust in human-machine partnerships and then building trust in the machine algorithms themselves. Second, trust needs to be established in relation to the amount of error or risk that an algorithm is allowed to accept. Autonomous technology is also a system of risk. "The key factor to assess with autonomous systems isn't whether the system is better than a human, but rather if the system fails (which it inevitably will), what is the amount of damage it could cause, and can we live with that risk" (193)? Third, trust figures into relegating the ethical and moral responsibility of warfare away from human agents and onto autonomous technologies. It is important to note that these processes are biopolitical and that the conversation about automation only addresses the side firing the guns. The victims of warfare remain vulnerable and also human.

12:15–12:30 p.m. **Closing Session**

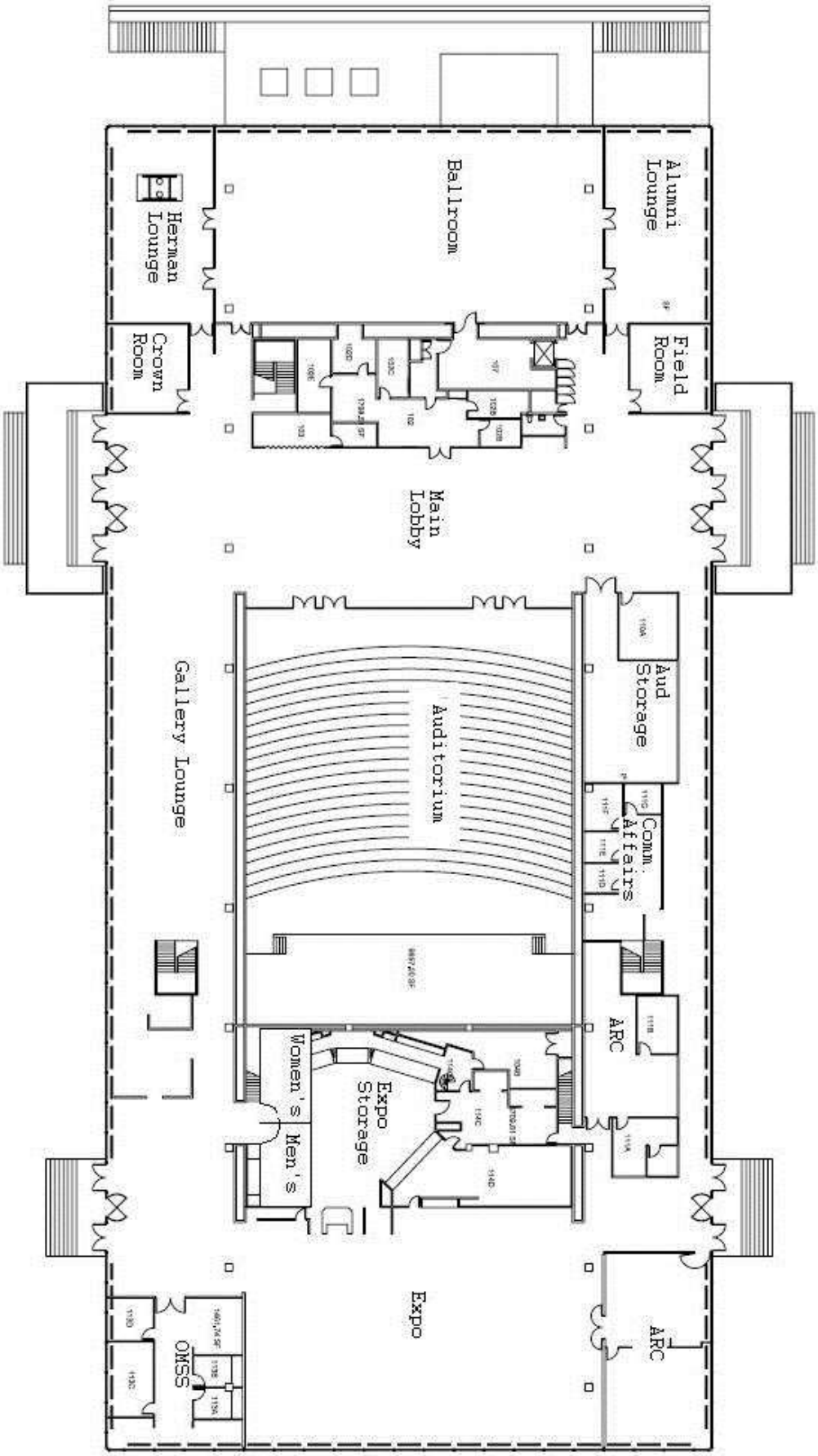
CEPE 2023 Conference Venue

Hermann Hall

3241 S Federal St
Illinois Institute of Technology
Chicago, IL 60616

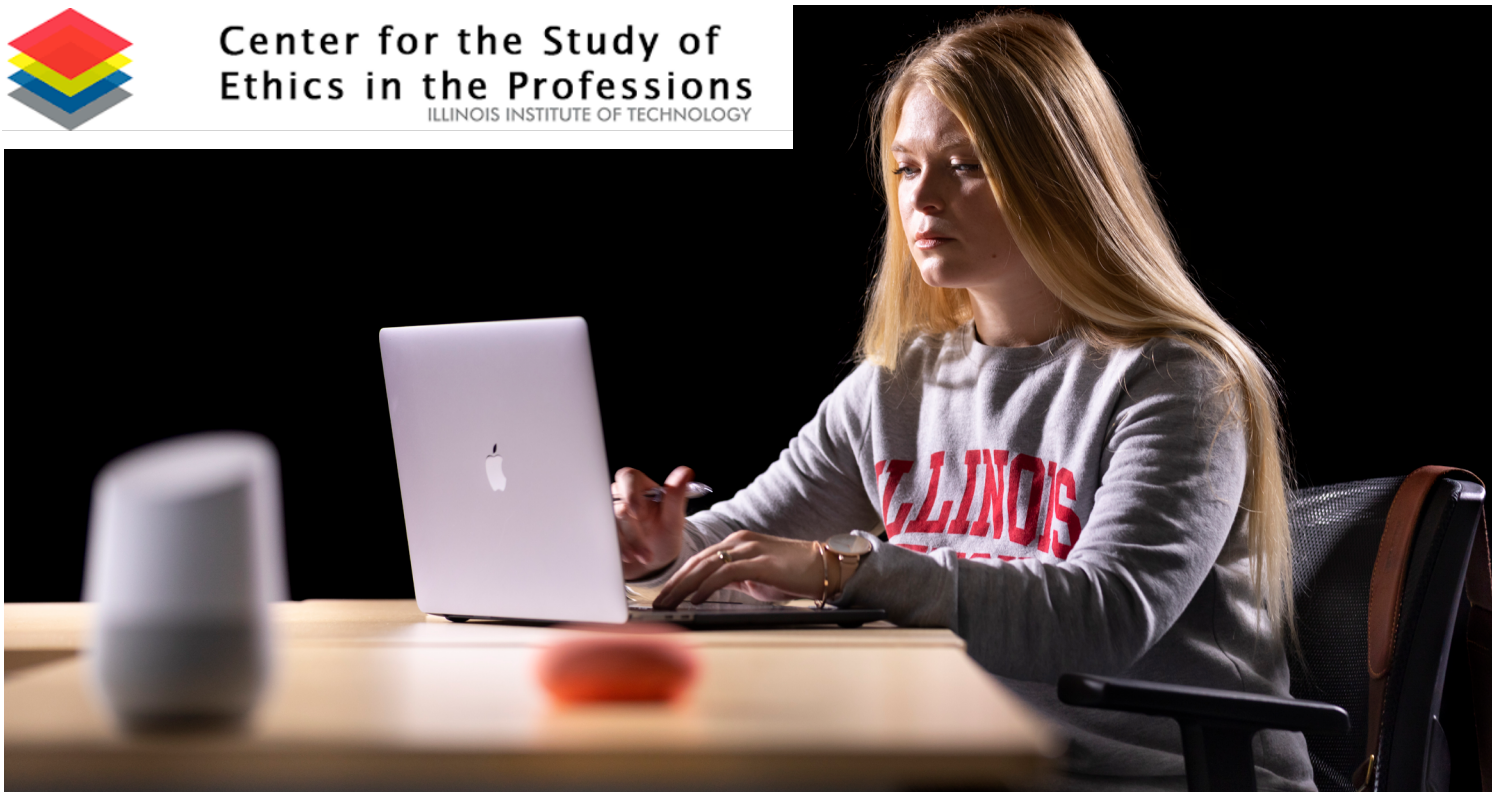


Map of Herman Hall (first floor)



South

Center for the Study of Ethics in the Professions
ILLINOIS INSTITUTE OF TECHNOLOGY



MISSION:

To educate students as responsible professionals, to reflect on the wider implications of scientific progress, and to contribute to the shaping of technology in accordance with fundamental human values.

VISION:

CSEP will be an internationally connected ethics center with a focus on professional and applied ethics, integrating ethics education into the colleges and departments of Illinois Tech, engaging in research and public dialogue at a local and global level.

WE SEEK TO:

- Enhance the distinctive education offered to Illinois Tech students by working with faculty from all the different colleges and departments at illnois Tech to help meaningfully integrate ethics into their educational programs – from the undergraduate to the graduate level.
- Promote innovative teaching by developing new pedagogical approaches and content in a wide variety of formats from the semester-long ethics course to shorter lessons, workshops or other formats. Establish a strong research program in ethics in the life sciences and in ethical and societal issues of emerging technologies.
- Build on the already existing unique CSEP collection of codes of ethics, expand and internationalize the collection to make it the basis for future research on codes of ethics.
- Be a strong participant in debates on ethical and societal implications of science and technology in the Chicago area, nationwide and internationally.

The Center for the Study of Ethics in the Profession's research program focuses on ethics in the life sciences and ethical and societal issues of emerging technologies, with a particular focus on philosophical and ethical aspects of neuroscience. The Ethics Center is committed to multi-disciplinary and multi-institutional research, to projects that combine empirical investigation with conceptual analysis, and to projects that introduce and propagate innovations in teaching. Furthermore, the Ethics Center Library houses a unique collection of ethics codes from all over the world and a large collection of ethics education materials.

Externally funded projects enable CSEP to conduct interdisciplinary research involving practitioners, as well as academics from Illinois Tech and other institutions. Topics CSEP has addressed include ways in which the brain and behavioral sciences might provide insight into moral and philosophical questions, intellectual property protection for science and technology, national security restrictions on the dissemination of scientific and technical information, responsible research and innovation in science, university/industry research relationships, organizational development, ethics in vocational education, and individual and collective responsibility in engineering.

Local Conference Organizers: Elisabeth Hildt (ehildt@iit.edu), Kelly Laas (laas@iit.edu)

ILLINOIS TECH