Suggestions for Final Essay Topics
Philosophy and AI 2024/2025

This document presents a list of essay topics that span epistemology, ethics, cognition, and the societal impacts of artificial intelligence. In each section, you'll find a brief description of the issues along with some literature that you may consult.

This list is **not exhaustive**. You are encouraged to explore these and other topics further and to propose your own ideas as discussed in lectures and tutorials.

Some of the guided questions below are quite general. For the essay, try to formulate a precise question (think of a concrete case or a specific aspect of a theory) as a starting point for your essay.

The Turing Test and the Nature of Intelligence

Over the decades, various tests have been used to measure machine intelligence. The Turing test, for example, evaluates a machine's ability to mimic human conversation, yet it's often criticized for achieving superficial imitation rather than genuine understanding.

- Is the Turing Test a valid measure of intelligence? What are the alternatives to the Turing Test for evaluating AI systems?
- How do alternative assessments (e.g., the Winograd Schema Challenge or the Minimum Intelligent Signal Test) address the limitations of the Turing Test?
- How might emerging evaluation methods be designed to better assess the full complexity of machine intelligence?
- What are the broader ethical and societal implications of relying on these tests to guide AI research and deployment?

References:

- Chollet, F. (2019). On the measure of intelligence.

The Singularity and Existential Risk

The technological singularity refers to a hypothetical future moment when artificial intelligence advances so rapidly that it surpasses human intelligence, triggering unprecedented changes across society, potentially leading to outcomes that are unpredictable and beyond human control.

- What defines, theoretically, the technological singularity?
- How do ethical and control challenges factor into the development of superintelligent AI?
- How might the integration of superintelligent systems reshape human identity and society?

References:

- Nick Bostrom – *Superintelligence: Paths, Dangers, Strategies* (2014)
- Adam Bales, William D'Alessandro, & Cameron Domenico Kirk-Giannini *Artificial Intelligence: Arguments for Catastrophic Risk* (2024, arXiv)
- Atoosa Kasirzadeh – *Two Types of AI Existential Risk: Decisive and Accumulative* (2024, arXiv)
- https://80000hours.org/problem-profiles/artificial-intelligence/

AI, Art, Beauty and Human Creativity

AI is reshaping the creative landscape by challenging traditional notions of art and beauty while also introducing controversial technologies like deepfakes. AI models generate striking images and democratize the artistic process, but deepfake algorithms can convincingly mimic human faces and voices.

- In what ways can AI enhance human creativity rather than replace it?
- What ethical dilemmas arise from using AI to generate art and deepfakes, particularly regarding consent and copyright?
- Can AI-generated art evoke the same emotional depth and human connection as art created solely by humans?
- What role does intention play in the creative process when AI tools are involved?
- What are the legal implications of using someone's likeness in deepfakes without their permission?
- How do societal perceptions of truth and trust change in an era where digital replicas and deepfakes are widespread?

References:

- Matthews, T., & Kidd, I. J. (2023). The Ethics and Epistemology of Deepfakes. *The Routledge Handbook of Philosophy and Media Ethics*, 342-354.
- "How AI can empower creators while respecting their rights", Vox https://www.vox.com/ad/385098/how-ai-can-empower-creators-adobe-firefly
- Hullman, Jessica, Ari Holtzman, and Andrew Gelman. "Artificial Intelligence and Aesthetic Judgment." (2023). https://arxiv.org/abs/2309.12338
- Chiodo, S., 2024. What AI "art" can teach us about art. *Journal of Aesthetics & Culture*, *16*(1)
- Atencia-Linares, P., & Artiga, M. (2022). Deepfakes, shallow epistemic graves: On the epistemic robustness of photography and videos in the era of deepfakes. *Synthese*, *200*(6), 518.
- Moruzzi, C. (2022). Creative agents: rethinking agency and creativity in human and artificial systems. *Journal of Aesthetics and Phenomenology*, *9*(2), 245-268.
- Paul, Elliot Samuel and Dustin Stokes, 2021, "Computer Creativity is a Matter of Agency", *Institute of Arts and Ideas News*, 11 November 2021

The Epistemology of Big Data

How is knowledge generated in an era defined by unprecedented volumes of information. Traditionally, scientific inquiry relied on theory-driven methods that emphasized causality and hypothesis testing; however, the advent of big data challenges this framework by privileging correlations over well-established causal models. Vast datasets are increasingly used to reveal patterns that seem to speak for themselves. Can raw data alone yield genuine understanding, or does the reliance on computational algorithms obscure the interpretive human element essential to knowledge?

References:

- Anderson, C. (2008). *"The End of Theory: The Data Deluge Makes the Scientific Method Obsolete."* Wired Magazine. URL

- Schönberger, V. M., and K. Cukier (2013). *Big Data. A Revolution That Will Transform How We Live, Work and Think.* London: John Murray.

- Boyd, D., & Crawford, K. (2012). *"Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon."* Information, Communication & Society, 15(5), 662-679.

- Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., & Suri, V. R. (2015). *"Big Data, Bigger Dilemmas: A Critical Review."* Journal of the Association for Information Science and Technology, 66(8), 1523-1545.

- Balazka, D., & Rodighiero, D. (2020). Big data and the little big bang: an epistemological (R)evolution. *Frontiers in big Data*, *3*, 31.

- Leonelli, Sabina, "Scientific Research and Big Data", *The Stanford Encyclopedia of Philosophy* (2020), Edward N. Zalta (ed) https://plato.stanford.edu/entries/science-big-data/

Ethics & AI

With the rapid development of AI, ethical questions regarding AI are becoming increasingly important. One key issue is whether it is possible to design autonomous agents that exhibit ethical behavior. Should specific rules be implemented, and which ethical framework (e.g., Kantian or Utilitarian) should be chosen? This question represents a clear top-down approach to the problem, but a bottom-up or hybrid approach is also possible.

Another crucial issue is the (future) ethical consequences of AI. Consider economic effects, such as job loss and increasing inequality. The rise of autonomous weapons and the ability to create deepfake news are also significant concerns.

The ethical implications of Weak AI and Big Data are additional points of interest. These include privacy concerns, the formation of echo chambers, and algorithmic biases that may have harmful or even discriminatory effects.

Another aspect is the potential dominance of major tech companies. These corporations have more resources (computing power) than universities, giving them a strong advantage in contemporary machine learning research. Will Google become the new Oxford, and is this desirable?

References:

- Bostrom, N. & Yudkowsky, E. (2011). *"The Ethics of Artificial Intelligence."* In *Cambridge Handbook of Artificial Intelligence*, W. Ramsey & K. Frankish (eds.). Cambridge: Cambridge University Press.

- Noorman, M. (2023). *"Computing and Moral Responsibility."* The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.). URL

- Müller, Vincent C., "Ethics of Artificial Intelligence and Robotics", *The Stanford Encyclopedia of Philosophy (2020)*, Edward N. Zalta & Uri Nodelman (eds.), https://plato.stanford.edu/entries/ethics-ai/

- https://iep.utm.edu/ethics-of-artificial-intelligence/

Some important journals:

- Science and Engineering Ethics

- Ethics and Information Technology

- AI and Ethics

A collection of recent papers:

https://philpapers.org/browse/ethics-of-artificial-intelligence

Relevant websites and organizations:

- https://futureoflife.org/

- https://www.humanetech.com/

Computation as a Conceptual Analysis of Cognition

Are computational processes sufficient for human-like cognition. Debates here include the limitations of computation compared to the complexities of human thought, the role of embodied cognition, and how modern AI developments increasingly mimic cognitive processes.

- What is meant by "computation" in the context of cognition? How do we define computation when applied to mental processes, and what distinctions exist between symbolic, connectionist, and dynamical approaches?

- Can computational models fully capture human cognition? What aspects of thought-such as creativity, consciousness, or embodied experience-might challenge a purely computational explanation?

- What role do mental representations play in computational theories? How do theories like the language of thought hypothesis or distributed representations account for the systematicity and productivity of human thought?

References:

- Van Gelder, T. (1995), *"What Might Cognition Be If Not Computation?"* URL

- Dreyfus, H. (1992), *"What Computers Still Can't Do: A Critique of Artificial Reason."* URL

- Sun, R. (2008). *Computational Cognitive Modeling*

- The Cambridge Handbook of Computational Cognitive Sciences (2023)
  https://doi.org/10.1017/9781108755610

- Finley, Kate. (2025). "Embodied cognition and the grip of computational metaphors."

- Chalmers, D. J. (2023). "The computational and the representational language-of-thought hypotheses"

- Stanford Encyclopedia of Philosophy: The Frame Problem

- [Stanford Encyclopedia of Philosophy: Embodied Cognition](#)

The Epistemology of ChatGPT

Do AI systems like ChatGPT "know" things? Is their knowledge akin to human knowing? Consider distinctions among knowing that, knowing how, and knowing why. How does interacting with such models affect human epistemic practices-possibly through the lens of the extended mind theory?

References:

- Clark, A. & Chalmers, D. (1998). *The Extended Mind*.

- Heersmink, Richard, et al. "A phenomenology and epistemology of large language models: transparency, trust, and trustworthiness." *Ethics and Information Technology*

- Helliwell, A. C. (2019). *Can AI mind be extended?*.

- [Stanford Encyclopedia of Philosophy: Analysis of Knowledge](#)

- [Stanford Encyclopedia of Philosophy: Content Externalism](#)

AI & Society

AI's is rapidly becoming a transformative force in our society, affecting nearly every aspect of our lives.

- What are the positive and negative consequences of AI integration on daily life and public policy? (think of a concrete application)

- How might AI-driven automation reshape the job market, and what strategies can help mitigate potential job displacement? Could AI lead to greater economic inequality, or does it offer opportunities to reduce the digital divide?

- What role does AI play in influencing public opinion, from spreading misinformation to countering conspiracy theories? How should society balance AI's benefits in enhancing communication with the risks of undermining democratic processes?

- How can we design AI systems that complement human creativity and decision-making without replacing the human element? What lessons from past technological revolutions (like the shift from horses to automobiles) can inform our approach to integrating AI into society?

References:

- Amodei, Dario, et al. (2016) "Concrete problems in AI safety".

- Whittlestone, J., & Clarke, S. (2022). "AI challenges for society and ethics"

The Ethics of AI in Education

Consider the implications of deploying AI (such as ChatGPT) in educational contexts. What ethical challenges arise concerning academic integrity, bias in educational algorithms, and the potential transformation of learning experiences? Evaluate both the potential benefits and drawbacks for educators and students.

- What are the ethical implications of using AI tools like ChatGPT in classrooms? (data privacy, misuse, ...) In what ways can bias in AI-powered educational algorithms affect learning outcomes?

- How does the integration of AI challenge traditional concepts of academic integrity? What strategies could be employed to ensure that AI's deployment in education remains ethical and equitable?

References:

- International Journal of Artificial Intelligence in Education

- https://www.theguardian.com/technology/2024/oct/05/chatgpt-has-become-the-best-teammate-to-these-sydney-university-students-but-is-there-a-limit

- https://www.myjournalcourier.com/news/article/teach-plus-calls-illinois-government-ai-policy-20136318.php (check the report mentioned in the article)

The Ontology of Artificial Consciousness

What does it mean for an AI system to be "conscious"? Can a machine possess subjective experiences, those intrinsic, qualitative "what-it-is-like" states that characterize human consciousness, and if so, how do such experiences differ from our own?

- How do various theories (such as functionalism, global workspace theory, and integrated information theory) explain the emergence of consciousness, and what implications do these explanations have for artificial systems? Can computational processes alone give rise to consciousness, or is a biological substrate necessary?

References:

- Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory of Consciousness*.

- Baars, B. J. (1988). *A Cognitive Theory of Consciousness*.

- Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). *Integrated Information Theory: From Consciousness to its Physical Substrate*.

- Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., ... & VanRullen, R. (2023). Consciousness in artificial intelligence: insights from the science of consciousness. *arXiv*

- Nagel, T. (1974). *What Is It Like to Be a Bat?*. The Philosophical Review

- Stanford Encyclopedia of Philosophy: Consciousness