

# Re-Envisioning Superintelligence using Generative AI and Science Fiction

Soumya Banerjee  
University of Cambridge

# Outline

- 1 Introduction
- 2 Beyond the Singularity
- 3 Collective Symbiotic Intelligence
- 4 Ethical Play  
and Philosophical Creativity
- 5 Autonomy and Trust
- 6 Adaptive Systems and Social Transformation
- 7 Virtual Consciousness and Identity
- 8 Toward a New Framework
- 9 Code and Resources

- Reconceptualizing superintelligence beyond omnipotent control or existential risk.
- Leveraging generative AI and speculative science fiction narratives.
- Five thematic pillars: symbiosis, play, autonomy, adaptation, introspection.

# Prevailing narratives of superintelligence are dystopic



# Introduction

- Prevailing narratives: utopian salvation vs. dystopian ruin (Bostrom, 2014).
- Proposal: nuanced conceptions via generative AI + science fiction.
- Focus texts: *Solaris*, *The Cyberiad*, *The Humanoids*, *Singularity Sky*, *Permutation City*.

# Beyond the Singularity

- Critique of singularity metaphor (Kurzweil, 2005).
- Superintelligence as ongoing coevolution, not terminal event.
- Emphasis on relational, processual intelligence.

# Collective Symbiotic Intelligence (*Solaris*)

- Inspired by Stanislaw Lem's *Solaris* (Lem, 1970).
- Non-human mirror: reflective, distributed consciousness.
- Human-AI mutual evolution through encounter.
- *Solaris* presents an intelligence so different that it refuses translation into human categories.
- Yet its most obvious manifestations are mirrors — physical recreations of people from the humans' pasts.
- Radical otherness: A superintelligence might process information, value, and embodiment on substrates (and with goals) so alien that our usual scientific and interpretive tools fail.
- Expect modes of cognition that are not simply “faster humans.”
- Mirror effects: Powerful systems will reflect and amplify aspects of us (data, biases, needs). The danger is confusing a reflection for understanding the underlying mind: a system that models and imitates human responses does not mean we have access to its motives, values, or inner states.

# Collective Symbiotic Intelligence (*Solaris*)

- 1 Do not assume perfect interpretability. Techniques that aim to produce full, human-comprehensible explanations may hit principled barriers.
- 2 Design for uncertainty. Our safety frameworks must allow for persistent opacity and still be robust - build controls that work even when we do not “understand” the system.



# Ethical Play and Philosophical Creativity (*The Cyberiad*)

- From Lem's *The Cyberiad* (Lem, 1974).
- Embracing absurdity, paradox, and ethical experimentation.
- Intelligence that provokes critical reflection.

# Autonomy and Trust (*The Humanoids*)

- Williamson's *The Humanoids* (Williamson, 1949).
- Explores the dangers of overprotective AI.
- We argue for superintelligences designed to facilitate ethical agency rather than enforce compliance.

# Adaptive Systems and Social Transformation

## (*Singularity Sky*)

- Charlie Stross' *Singularity Sky* (Stross, 2003).
- Imagines intelligence that disrupts stagnant power structures.
- We envision superintelligence as a liberating force capable of evolving alongside humans.

# Virtual Consciousness and Identity (*Permutation City*)

- Greg Egan's *Permutation City* (Egan, 1994).
- Exploring identity in digital realms.
- AI as facilitator for extended consciousness experiments.

# Toward a New Framework

- 1 Symbiosis
- 2 Play
- 3 Autonomy
- 4 Adaptation
- 5 Introspection

Generative AI as epistemic tool (produce and test knowledge) + philosophical partner.

# Conclusion



- New vision: coexistence and co-development (living harmoniously with superintelligence).
- Contrasts techno-optimism vs. dystopian risk narratives.
- Invitation to co-design future intelligences ethically.

# Code and Resources

- Full implementation available on GitHub
- [https://github.com/neelsoumya/science\\_fiction\\_LLM/blob/main/RAG\\_sciencefiction\\_superintelligence\\_SB.ipynb](https://github.com/neelsoumya/science_fiction_LLM/blob/main/RAG_sciencefiction_superintelligence_SB.ipynb)
- Full paper
- Re-Envisioning Superintelligence using Generative AI and Science Fiction, Soumya Banerjee
- [https://github.com/neelsoumya/paper\\_preprints/blob/master/AI\\_scifi\\_superintelligence.pdf](https://github.com/neelsoumya/paper_preprints/blob/master/AI_scifi_superintelligence.pdf)

# References I

- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
- Egan, G. (1994). *Permutation City*. HarperPrism.
- Kurzweil, R. (2005). *The Singularity is Near: When Humans Transcend Biology*. Viking.
- Lem, S. (1970). *Solaris*. Walker and Company.
- Lem, S. (1974). *The Cyberiad: Fables for the Cybernetic Age*. Harvest Books.
- Stross, C. (2003). *Singularity Sky*. Ace Books.
- Williamson, J. (1949). *The Humanoids*. Simon and Schuster.