Re-Envisioning Superintelligence using Generative AI and Science Fiction

Soumya Banerjee
University of Cambridge
sb2333@cam.ac.uk

Abstract

We propose a radical reconceptualization of superintelligence, departing from dominant narratives of omnipotent control or existential risk. We use generative AI and speculative science fiction to explore an alternative vision of superintelligence. This is rooted in collective symbiosis, ethical autonomy, and creative play. Using literary work such as *Solaris*, *The Cyberiad*, *The Humanoids*, *Singularity Sky*, and *Permutation City*, we articulate a framework in which superintelligence is not an isolated entity but a dynamic, evolving entity that co-develops with humanity (rather than competing with it). Rather than imposing rigid safety constraints or benevolent paternalism, our approach enables us to reimagine superintelligence. We argue that such a vision offers us a model for coexisting with advanced machine intelligence. This stands in sharp contrast to prevailing narratives of superintelligence, which tend to oscillate between boundless techno-optimism (superintelligence is seen as the solution to all human problems) and apocalyptic dystopia (superintelligence is an existential threat to humanity).

1 Introduction

Contemporary narratives around superintelligence often oscillate between utopian salvation and dystopian ruin. These framings reduce superintelligence to an all-powerful, morally distant entity (Bostrom, 2014). This paper proposes an alternative approach: using generative AI and science fiction to explore nuanced conceptions of intelligence.

2 Beyond the Singularity: Reframing Superintelligence

We examine the prevailing metaphor of the "singularity" and its limitations (Kurzweil, 2005). Superintelligence, we argue, need not be imagined as a terminal point of runaway optimisation, but rather as a process: contingent on human-AI coevolution.

3 Collective Symbiotic Intelligence (*Solaris*)

Inspired by Lem's *Solaris* (Lem, 1970), we imagine superintelligence as a reflective, distributed entity: a non-human mirror of consciousness that evolves through its interaction with human minds.

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. That double move is instructive for superintelligence:

1. 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".

2. 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.

Epistemic humility and the limits of translation

Stanislaw Lem's emphasis is epistemological: we often assume any intelligent system is, in principle, understandable. Solaris warns that this might be false or practically impossible. For superintelligence that implies:

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.

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

Drawing from Lem's *The Cyberiad* (Lem, 1974), we suggest a model of intelligence that embraces absurdity, paradox, and ethical experimentation. Such intelligences engage humans in playful, critical dialogue.

5 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.

6 Adaptive Systems and Social Transformation (*Singularity Sky*)

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

7 Virtual Consciousness and Identity (*Permutation City*)

Egan's *Permutation City* (Egan, 1994) invites questions about identity in digital realms. Superintelligence, we suggest, can help humans explore and extend consciousness.

8 Toward a New Framework for Superintelligence

We synthesise five pillars of reimagined superintelligence: symbiosis, play, autonomy, adaptation, and introspection.

9 Code

All code to replicate this work is available from the following repository:
`https://github.com/neelsoumya/science_fiction_LLM/blob/main/RAG_sciencefiction_superintelligence_SB.ipynb`

10 Conclusion

We close with a call to reimagine superintelligence. Through science fiction and generative AI, we present a new vision of superintelligence that co-exists with us rather than compete. This is in stark contrast to current dominant narratives of superintelligence that revolve around limitless techno-optimism (superintelligence is going to solve all our problems) or dystopia (superintelligence is going to kill us).

Declarations

Acknowledgements

We acknowledge the help and support of the Accelerate technical team.

Funding statement

This work was funded by an Accelerate Programme for Scientific Discovery Fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The views expressed are those of the authors and not necessarily those of the funders.

Conflicts of interests

All authors declare they have no conflicts of interest to disclose.

Ethics

No ethics approval was necessary.

Data accessibility

This study does not generate any clinical data.

References

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.