Complementary Artificial Intelligence

"Artificial Intelligence" was coined by McCarthy in 1955 to headline a workshop that canonized Turing's "imitation game" approach to mimic human intelligence[1]. This notion of AI has dominated computational intelligence designs for more than half a century but is not tuned to accelerate humanity's advance against its greatest challenges by complementing human capacity. With 8 billion humans on earth and stagnating labor productivity growth, halving to 1.3% for all but one OECD country since the 1990s[2], we argue that the production of imitating artificial intelligence is neither society's most strategic nor ethical computational investment.

This vision contrasts with the "Augmented"[3] Intelligence vision put forward by Englebart and others. Associated with the rise of personal computing, this approach sought to accelerate thinking by enabling unmediated human access to information for reference and manipulation. Of the two AIs, augmented intelligence has been more commercially successful, but is necessarily incremental as tethered to human capacity. For example, in science, technology, and medicine, with millions of professionals competing against each other and nature, would the better AI marginally increase human capacity or augment it by avoiding its blind spots and overcoming its limitations?

We reconceptualize beneficial AI as computational intelligence that radically complements human understanding by thinking differently, expanding human cognitive capacity rather than competing with it. A complementary approach to AI relies on the phenomenon that diverse ensembles dominate. The 'wisdom of crowds', where groups outperform individuals, hinges on the independence and diversity of information[4] and approach[5] held by its members. The wisdom of crowds is also manifest in algorithmic crowds: Data science competitions like Kaggle where ensemble models (e.g., boosted forests or deep networks) have always bested singular models. With more data and better models than ever before on how individual and collective intelligence operates in the world, we are now uniquely in a position to cultivate AI with novel, complementary objectives, which avoid human overthinking to identify promising leads inconceivable without it.

These complementary AI could take on roles of human coach, collaborator, and coordinator. Complementary AI could coach individual human limitations and cognitive bias. Judges demonstrate systemic bias in bail hearings, which can be discovered and offered constructive feedback from algorithms encoding shared values like fairness[6]. High-performance game-playing AI, designed to beat human opponents, unintentionally inspires novel human performance[7]. AI could more directly complement human capacity. Complementary AI could also compensate for collective human limitations embedded within social and technical institutions. We have demonstrated an approach to complementary AI by designing algorithms that productively *collaborate* with human researchers who study energy-related and pharmacological materials by avoiding what human scientists are likely to study alone or in concert, dramatically improving total performance[8]. In contexts of civil discourse or corporate problem-solving, we could similarly design AI collaborators with generative language models to burst group-think or polarized deliberation. Alternatively, consider how the historical under-representation of black patients to medical professionals has likely led to the profession of radiology being less able to identify black than white patient joint pain from color-blind X-rays[9]. Complementary AI could *coordinate* algorithms with properties and service providers with experiences that collectively maximize accurate

diagnoses and helpful treatments. Building human-aware AI coaches, collaborators, and coordinators to avoid the crowd could generate complementary insights that accelerate scientific discovery, medical treatment, and policy compromise by designing the collective diversity humans and algorithms require to succeed together.

Building complementary AI raises challenging technical and ethical issues. These include the paradox of how to build and guide algorithms to think things we cannot think ourselves. Insofar as individual and collective human intelligence has been refined over evolutionary eons for performance on a wide range of existential problems, designing AI that simply thinks differently will produce low-quality intelligence. We need to incorporate additional signals that indicate where human intelligence is maladaptive to evolve complementary AI. With increasing AI performance and autonomy, scholars raise cogent concerns regarding AI alignment with human values[10], but the logic of control typically recommended to discipline AI must be tempered with the alternative logic of care, where caregivers "raise" those cared for to discover new, beneficial capacities that contribute to the collective in ways which cannot be fully anticipated[11].

We propose a systematic, AI-enabled research program to discover and program beyond individual and collective human limitations and design complementary AI that assists humanity to become more intelligent, creative, and productive together[12]. We propose an associated research design in the Figure. The behavioural, social, and cultural sciences will be of critical importance in concert with computer scientists, mathematicians and engineers in helping AI realize its capacity to radically complement human capacity, unleashing new potential, productivity and promise.

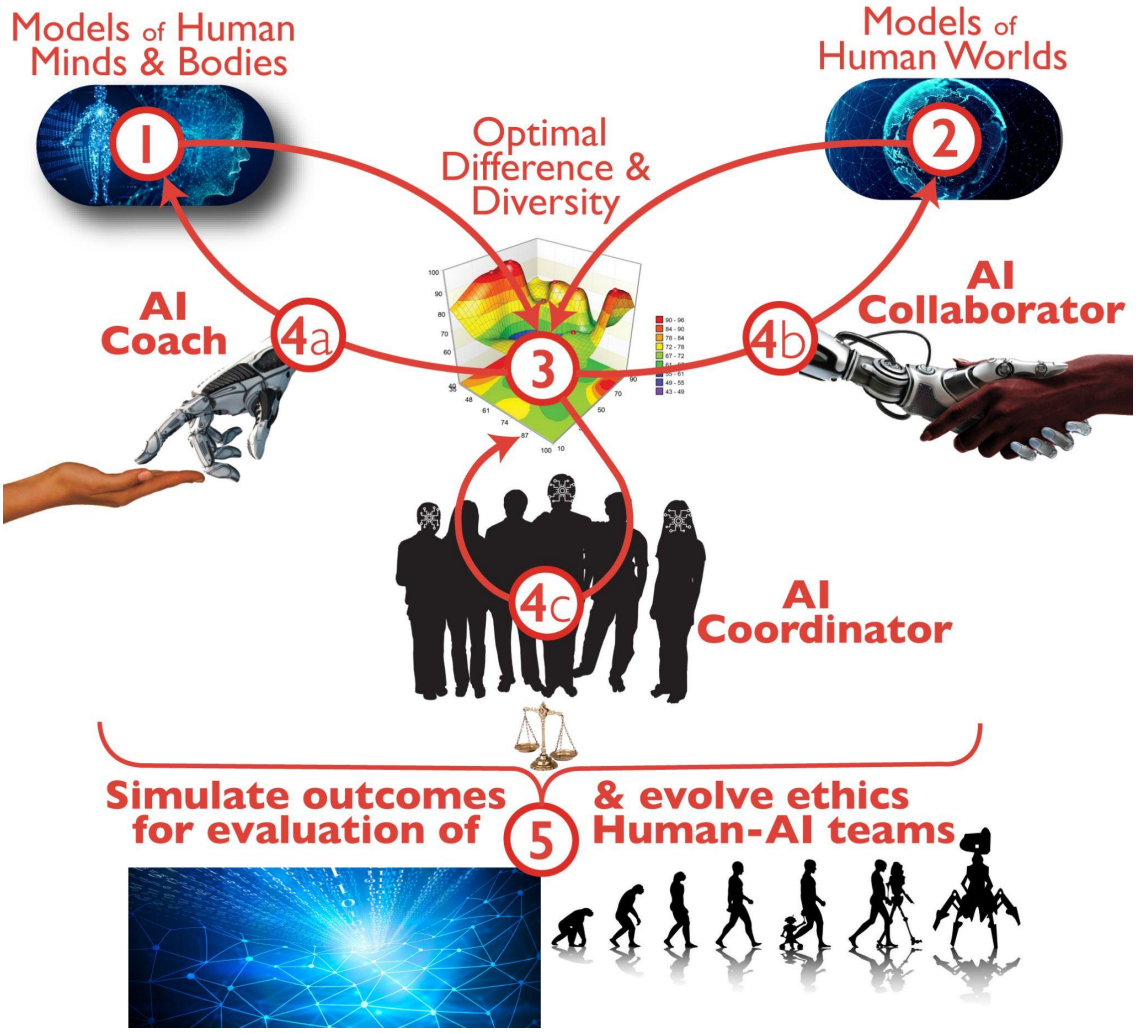

Fig. A research program that leverages models human minds and bodies operating in human worlds to optimally evolve complementary AI that coaches, collaborates with, and coordinates their human and machine counterparts; evolving and simulating the consequences of configurations, and exploring a new ethics that accounts not only for a paradigm of control, but also care-giving.

Jamshid Sourati[a], Chenhao Tan[a], and James Evans[a,b*]

[a]University of Chicago, 1155 S. 60th Street, Chicago, IL 60637
[b]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501

References

1. Turing, A. M. I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind* vol. LIX 433–460

   Preprint at https://doi.org/10.1093/mind/lix.236.433 (1950).

2. Brynjolfsson, E., Rock, D. & Syverson, C. *Artificial Intelligence and the Modern Productivity*

Paradox: A Clash of Expectations and Statistics. https://www.nber.org/papers/w24001 (2017) doi:10.3386/w24001.

3. Englebart, D. E. Augmenting human intellect. *originally published in' The Augmentation Papers', Bootstrap Institute* 64–90 (1962).
4. Surowiecki, J. The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business. *Economies, Societies and Nations* **296**, (2004).
5. Page, S. E. *The Diversity Bonus: How Great Teams Pay Off in the Knowledge Economy*. (Princeton University Press, 2019).
6. Ludwig, J. & Mullainathan, S. Algorithmic Behavioral Science: Machine Learning as a Tool for Scientific Discovery. (2022) doi:10.2139/ssrn.4164272.
7. Shin, M., Kim, J., van Opheusden, B. & Griffiths, T. L. Superhuman artificial intelligence can improve human decision-making by increasing novelty. *Proceedings of the National Academy of Sciences* **120**, e2214840120 (2023).
8. Sourati, J. & Evans, J. Accelerating science with human versus alien artificial intelligences. *arXiv [cs.AI]* (2021).
9. Pierson, E., Cutler, D. M., Leskovec, J., Mullainathan, S. & Obermeyer, Z. An algorithmic approach to reducing unexplained pain disparities in underserved populations. *Nat. Med.* **27**, 136–140 (2021).
10. Bostrom, N. *Superintelligence: Paths, Dangers, Strategies*. (Oxford University Press, 2014).
11. Gopnik, A. *The Gardener and the Carpenter: What the New Science of Child Development Tells Us About the Relationship Between Parents and Children*. (Macmillan, 2016).
12. Wilder, B., Horvitz, E. & Kamar, E. Learning to Complement Humans. *arXiv [cs.AI]* (2020).