Toward a Framework for Cooperative Hybrid Intelligence

Rayne Wallace, Prabhdeep Singh, and Vihaan Sharma
October 26, 2025

Abstract

The advent of multimodal AI agents has greatly enhanced the integration of artificial intelligence with human life. By interacting through vision, text, and audio, these agents can automate a wide array of tasks that previously required human intelligence. However, many agent systems are task-specific, with interfaces optimized for explicit prompts of predefined utilities. Here we propose a framework for the radically extended integration of AI agents, moving away from the paradigm of AI as a tool, and toward a concept of cooperative hybrid intelligence (CHI). CHI has been explored in the context of co-evolutionary systems [1], social science [2], and education [3]; yet broad theoretical formulations of CHI remain limited. Here we show that the dynamics of collaborative intelligent entities can be modelled using existing mathematical frameworks of cognition and cooperation. In particular, we examine applications of the free energy principle [4, 5], and their consequences for cooperative optima between intelligent systems. This framework is intended to guide the development of future CHI systems, providing a rigorous theoretical foundation for the union of computer and man.

Introduction

The integration of multimodal AI agents into daily life has significantly transformed the role of artificial intelligence in enhancing human capabilities. These agents, capable of interacting through vision, text, and audio, enable automation of a broad range of tasks that once required human intelligence, from simple data processing to complex decision-making. However, despite their impressive capabilities, most current AI systems remain task-specific, designed with interfaces optimized for explicit prompts and predefined functions. These systems often operate in isolation, with limited ability to adapt to new contexts or engage in more fluid, dynamic interactions.

In this paper, we introduce a framework for a more expansive vision of AI integration—cooperative hybrid intelligence(CHI)—which moves beyond the traditional notion of AI as a mere tool. Rather than being confined to specific, isolated tasks, CHI envisions AI as a collaborative partner that works in concert with humans and other intelligent entities. While the concept of CHI has been explored in areas such as co-evolutionary systems [1], social science [2], and education [3], comprehensive theoretical frameworks that underpin these ideas remain relatively underdeveloped.

We argue that the dynamics of collaboration between intelligent entities, whether human or artificial, can be modelled using existing mathematical frameworks of cognition and cooperation. In particular, we explore the application of the free energy principle [4, 5] as a foundational model for understanding how cooperative optima might emerge between intelligent systems. By applying this principle, we aim to provide a rigorous and unified theoretical basis for understanding and guiding the development of future CHI systems. Ultimately, this framework aspires to support the evolution of more integrated and adaptive interactions between humans and artificial intelligence, fostering deeper and more effective collaboration across a range of domains.

Formalizing Cooperative Dynamics

To begin our analysis, it will be necessary to formalize the dynamics of cooperation between intelligent agents. This evidently depends quite closely on the concept of *agency*; here we borrow a definition of agency common in existing literature: agents are entities that act purposefully to achieve specific goals, and adapt to their environment in pursuit of those objectives [6, 7]. It follows quite naturally that agents must have an intrinsic measure of success or failure, that is, a reward metric, which serves as a proxy for concrete success in the given objectives. This is true for even the simplest objectives,

since the outcome can only be assessed via some sensory modality, which is necessarily a proxy for the actual occurence. It is noteworthy that this conception of agency is highly compatible with the existing paradigm of agentic machine learning [8, 9], which depends on explicit reward, loss, or regret metrics to optimize model performance.

With this in mind, we can consider each agent as an entity that optimizes a set of reward metrics as a function of context and policy. Formally, for an agent A, we have $R_A(\pi, \gamma_A)$, where π is a policy and γ_A is the agent context. In practical use, the relevant value is $\hat{R}_A(\pi, \gamma_A)$, the empirical reward for a given action consistent with π. Since other (cooperating) agents are external to A, we can consider the actions of another agent B to be a component of γ_A, and the expected impact of B on γ_A can be denoted $\Delta\gamma_A(B)$. We can now define the *policy synergy* of π for a given baseline context γ_A as:

$$s_\pi(B) = \frac{\mathrm{E}\left(\hat{R}_A(\pi, y_A + \Delta\gamma_A(B))\right)}{\mathrm{E}\left(\hat{R}_A(\pi, y_A)\right)}$$

It is now helpful to consider the distribution of $s_\pi(B)$ over all policies $\pi \in \Pi_A$. We can define $\bar{s}(B)$ as the mean policy synergy for A given B. Bear in mind that the reward metric for A may or may not align with the reward metric for B. Moreover, the policy with highest synergy may not align with the baseline optimal policy, which is $argmax_{\pi_i}\left(R_A(\pi_i, \gamma_A)\right)$. This synergetic misalignment can be expressed as the Kullback-Leibler divergence of two reward distributions: the baseline distribution Q, and P, the distribution of reward in the presence of B. This will give rise to some interesting dynamics, as we shall see shortly.

For more practical purposes, suppose that A is a machine learning model, while B is a human being. If A is optimally trained, it will follow the reward-optimal policy $argmax_{\pi_i}\left(R_A(\pi_i, \gamma_A)\right)$. Yet if $D_{KL}(P||Q)$ is high, A will experience significant loss while interacting with B. This is, by definition, a suboptimal cooperation. The same is true if the agents switch places, with the human in the place of A and vice versa. This poses two main challenges for the development of cooperative artificial intelligence.

Firstly, human beings have extensive experience interacting with their surroundings, and human action policies are heavily based on this experience. This experience also occurs in an ergodic system of contextual states. Consequently, humans are unlikely to radically change policy in a novel situation, if this policy change would come at a great loss in the baseline ergodic context-set. This notion is supported by the

psychological literature [10], and is consistent with widely accepted models of learned behaviour. In other words, humans cannot be expected to significantly alter learned policy simply to optimize interaction with a machine learning model. This is challenging if the presence of a cooperative intelligent agent significantly changes the reward distribution.

Second, artificial intelligence models are generally trained outside a human-embedded context. Training data is generally highly curated and focused on specific tasks. Thus, it may be a poor representation of the model's reward distribution in the presence of a human. This concern has been addressed somewhat more than the first, with methods such as reinforcement learning with human feedback (RLHF) providing more human-oriented representations of reward [11]. However, human feedback in a controlled training environment does not fully capture the full range of human action, since the model is exposed only to the specific training feedback, and not to the complete effect of the human on the environmental state. Thus, even a model trained using state-of-the-art techniques will generally experience surprisal when interacting closely with a human being.

With this in mind, we can define a critical training objective for cooperative hybrid intelligence, namely to minimize the Kullback-Leibler divergence between the policy-reward distributions in the baseline versus cooperative contexts. In other words, a well-trained model for CHI should minimize the surprisal, ie. variational free energy, of observed reward given a human-interactive context. Under this paradigm, the optimal policy is then:

$$argmin_{\pi_i}\left(-logP(\hat{R}\,|\,\gamma)\right)$$

This is far from trivial to train directly. It also must be balanced against a policy that is functional overall; it would not do to learn a policy with a highly accurate expectation of dismal reward across contexts.

As one approach, we propose *convergent optimization*, in which components of a hybrid intelligent system learn predictive models of one another. This reduces the surprisal of interaction for both parties, since each agent learns a distribution of the other's actions, and by extension, of $\Delta\gamma_A(B)$. This convergent modelling may also reduce cognitive friction for humans, since the model's behaviour will be perceived as more native and expected, reducing the human's surprisal. Convergent optimization could be implemented in many ways, including by directly mimicking certain aspects of the human's behaviour. For instance, by learning an imitation of a human collaborator's voice, a model may reduce sensory surprisal for both parties.

Finally, we consider the possibility of *altruistic optimization*, which acts to optimize shared reward instead of treating each agent as selfishly distinct. This poses some difficulty because both agents will necessarily have prior learning experiences outside the context of the cooperative partnership. Humans have years of life history, and machine learning models will presumably have undergone some training by the time they begin to cooperate with humans in an embedded context. Thus, each agent has individual objectives and learned reward distributions in addition to the shared objectives. Any attempt to avoid this, such as choosing not to pretrain the model at all, would lead to obvious problems, such as the model being utterly useless for a significant period of time at the start of the cooperation. However, good results may be found in a compromise between objectives, in which each agent makes some sacrifice in individual reward to contribute to shared reward. This shared reward will, in the beginning, not be so deeply engrained as the individual reward. Suppose that a given policy results in a sacrifice in individual reward of magnitude δ, and an increase in shared reward of equal magnitude. If R_s and R_A are the shared and individual reward, respectively, then A may give weight to the two objectives such that a given increase δ in R_s is worth only a $\lambda\delta$ increase in R_A for some scaling coefficient $\lambda < 1$. In this case, the loss expressed by λ must be compensated by the synergy between agents. Thus a cooperative policy is beneficial if

$$\frac{\lambda s_\pi(B)\delta_\pi}{\delta_\pi} > 1$$

Readers familiar with cooperative theories in biology will recognize the similarity to Hamilton's equation for kin selection [12]. This is quite fitting, since both situations involve cooperation between independent agents when shared reward cannot be forced externally. It may be of use to investigate further parallels between the mathematical behaviour of animals and artificial intelligence.

Conclusion

In conclusion, we show that the interactions between cooperating agents may be modelled using existing frameworks for minimizing variational free energy. This line of investigation suggests methods for training CHI systems that emphasize modelling of human collaborators. This strategy reduces surprisal for both agents, and may provide the foundation for future innovations in cooperative hybrid intelligence.

References

[1] Krinkin, K., Shichkina, Y., & Ignatyev, A. (n.d.). Co-evolutionary hybrid intelligence. *arXiv*.

[2] Akata, Z., Balliet, D., de Rijke, M., Dignum, F., Dignum, V., Eiben, G., Fokkens, A., Grossi, D., Hindriks, K., Hoos, H., Hung, H., Jonker, C., Monz, C., Neerincx, M., Oliehoek, F., Prakken, H., Schlobach, S., van der Gaag, L., van Harmelen, F., … Welling, M. (2020b). A research agenda for Hybrid Intelligence: Augmenting Human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. *Computer*, *53*(8), 18–28. https://doi.org/10.1109/mc.2020.2996587

[3] Kong, X., Fang, H., Chen, W., Xiao, J., & Zhang, M. (2025). Examining human–AI collaboration in Hybrid Intelligence Learning Environments: Insight from the synergy degree model. *Humanities and Social Sciences Communications*, *12*(1). https://doi.org/10.1057/s41599-025-05097-z

[4] Ramstead, M. J., Sakthivadivel, D. A., Heins, C., Koudahl, M., Millidge, B., Da Costa, L., Klein, B., & Friston, K. J. (2023). On bayesian mechanics: A physics of and by beliefs. *Interface Focus*, *13*(3). https://doi.org/10.1098/rsfs.2022.0029

[5] Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology*.

[6] Jaeger, J., Riedl, A., Djedovic, A., Vervaeke, J., & Walsh, D. (2024). Naturalizing relevance realization: Why agency and cognition are fundamentally not computational. *Frontiers in Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1362658

[7] Naidoo, M. (2023). What does it mean to be an agent? *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1273470

[8] Sutton, R. et. al.(2000). Policy Gradient Methods for Reinforcement Learning with Function Approximation. *Advances in Neural Information Processing Systems*.

[9] Silver, D. (n.d.). Deterministic Policy Gradient Algorithms. *DeepMind Technologies*.

[10] Wang, Y., Tian, J., & Yang, Q. (2024). Experiential Avoidance Process Model: A Review of the Mechanism for the Generation and Maintenance of Avoidance Behavior. *Psychiatry Clin. Psychopharmacol.*

[11] Christiano, P. (2017). Deep reinforcement learning from human preferences. *arXiv*.

[12] Hamilton, W. D. (1964). The Genetical Evolution of social behaviour. II. *Journal of Theoretical Biology*, *7*(1), 17–52. https://doi.org/10.1016/0022-5193(64)90039-6