A Physics and Neuroscience-Inspired AI Framework for a Sustainable Future

From Symbolic Models to Collective Intelligence

Thomas P. Kehler, Ph.D.

Abstract

This paper describes the evolution of a next-generation AI architecture that not only reduces risk and incorporates human cognition but also holds the potential to revolutionize the field. The new architecture incorporates fundamental principles from the natural sciences, specifically physics, computational biology, and neuroscience. The first iteration of this architecture combines the science of collective intelligence with an adaptive learning approach that builds knowledge models from the collaborative work of

human and AI agents working together. Early work on the application of developing causal models of investment decisions was published in a paper for NIPS 2018.[1] That work was extended to a general AI architecture for decision intelligence discussed here.

The work is set in the larger context of AI over the past five decades. The first wave of AI focused on creating symbolic models of human reasoning and knowledge representation. These models, based on logic and network models of semantics, were explanatory and contextualized. However, they were limited in their capacity for self-extension—they did not learn from data.

The current generation of AI, which applies statistical methods, has resulted in significant advances in deep learning and pre-trained transformer models. These models show a significant ability to scale and demonstrate impressive capabilities. However, because they are learning from complex aggregated statistical correlations, they lack explanatory capabilities and thus open the door to distrust and concerns about future impact on humanity. In addition, they need a clearer method of managing context. Training massive models with a wide context window is energetically challenging. They threaten struggling climate change initiatives, highlighting the urgent need for a new approach.

Over the past decade, new distributed AI models have emerged that promise a path forward that addresses explanatory power, context management, and energy efficiency issues. These models are grounded in physics and biology. Furthermore, the essence of the new AI architecture is to amplify the co-creative of collective intelligence. It is collective intelligence, institutionalized in the scientific method of knowledge discovery, that has led to all humanity's progress to date.

Key aspects of how the third wave of AI will be presented include a survey of results in applying the neuroscience-based work associated with the Free Energy Principle and Active Inference. In addition, results will be presented in the integration of the science of collective intelligence with new AI methods and the implications of these methods to multi-agent models that integrate human collective intelligence with LLM-based AI agents.

A 50-year perspective on Artificial Intelligence

Very early AI

In the summer of 1969, I began my journey with generative models and AI. The journey began in applied physics, and the problem was to infer what underlying magnetic

---

[1] Kehler, T., Liu, F., Olfat, M., & Sinha, S. (2024, April 8). Predicting Startup Funding Momentum with Collective Intelligence. https://doi.org/10.31219/osf.io/k6wq5

behaviors, as measured in the magnetic susceptibility χ, would generate certain data outputs in an experimental apparatus. A generative model of magnetic susceptibility χ was constructed that computed χ from a theoretical representation (causal) of the underlying physics. From this, one could compute the expected values in the experimental apparatus. A cost function comparing expected and measured values was minimized using gradient descent techniques to extract χ. From this, we learned that the then-current magnetization ripple theory failed to predict observed data. Modifying the theoretical model led to an improved prediction of χ.

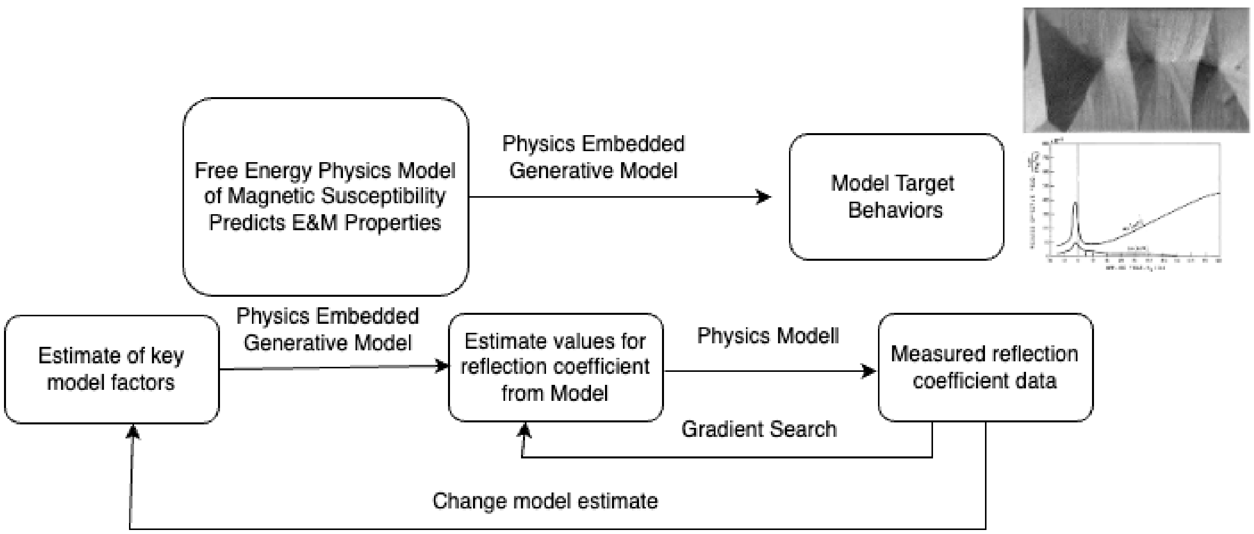


Figure 1

Computational support for machine learning came early in computing history. In 1969, the IBM Scientific Programming Library contained methods for cost function minimization, very similar to methods used today.

The approach is generative and based on a causal model grounded in experimental and theoretical physics. Early AI models used ‘generate and test’ as a paradigm.

Three waves of AI: Symbolic, Statistical, Contextual/Adaptive

Most early AI efforts derived from symbolic logic and symbolic representation of knowledge. Early techniques focused on hand-crafted models designed to mimic human problem-solving models.

The figure below captures a broad perspective on AI. It is from a DARPA Perspective on AI published over six years ago.

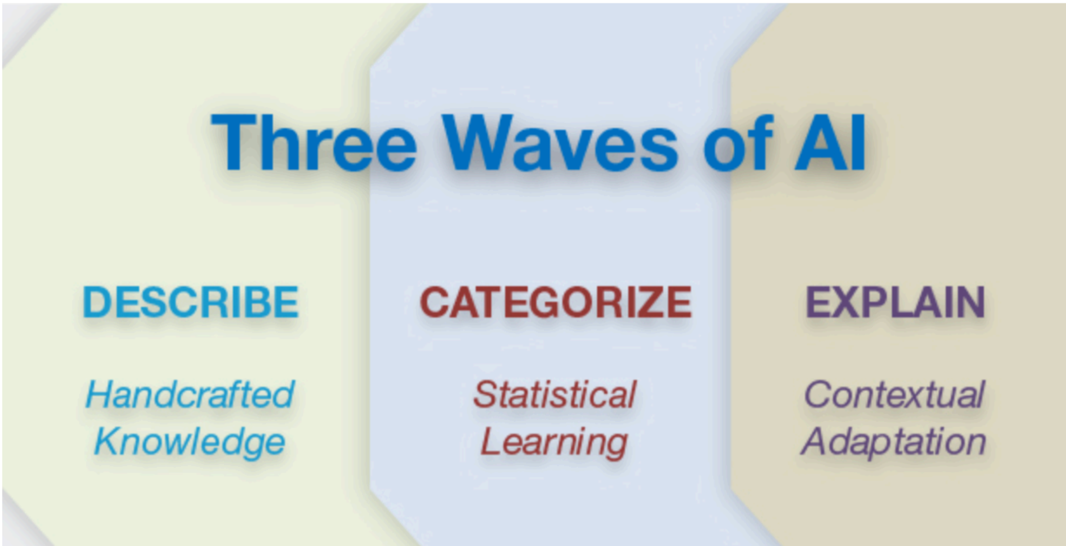


Figure 1

The first wave of AI occurred in the 1980s and was based on symbolic models. A key feature of the handcrafted knowledge models was their ability to explain. Systems were built on symbolic representations of knowledge and inference. I was a co-creator of one of the leading development platforms of that era: KEE - the Knowledge Engineering Environment.[2] Human knowledge was encoded in networks of symbolic entities and relationships (semantic networks) and reasoning systems (symbolic rule-based logic).[3]

The learning cycle to train these systems is manual. The knowledge modeling and reasoning process is extracted from interviews with experts and associated data and documents. Explanations from the system were part of the training process. The correction to the model was manual. For this reason, while these systems work for many decision automation tasks, they do not adaptively learn from data. We will return to this point when discussing an automated collective knowledge acquisition system.

Explanation is the bedrock of trust for human-AI interactions. Statistical learning is unique in that it does not provide explanations, and thus, it has struggled with the problem of trust. The ability of an AI to explain its thinking is critical to building trust and working with humans. The middle part is where we are now, so we struggle to trust it. Large language models will tell you they are not to be trusted but to use critical thinking. An example of this is shown in the query and response below:

2 Kehler, Thomas P., and Gregory D. Clemenson. "KEE the knowledge engineering environment for industry." *Systems And Software* 3.1 (1984): 212-224..

3 Richard Fikes and Tom Kehler. The role of frame-based representation in reasoning. Communications of the ACM, 28(9):904–920, 1985.

Question to ChatGPT: Do Large Language models provide a way to validate the source of knowledge?

ChatGPT: Large language models by themselves do not provide a way to validate the source of knowledge. These models are trained on vast amounts of text data and can generate text that is similar to the text they were trained on, but they do not have the ability to verify the accuracy or source of the information they provide. It's important for people to use critical thinking skills and verify information from multiple sources in order to determine its accuracy and credibility.
In contrast, the Meta-Agent creates a complete audit trail of how it constructed the knowledge model of human collaboration so that the knowledge model it produces can explain how it gets its answers. We also require it to keep track of the provenance of that information.
This AI does not care if the agents are human or AI Agents. We can then quickly see that this 'Meta-Agent' can empower humans to curate the generative outputs of AI Agents based on LLMs or other AI technologies. This later property of curation enables human participants to apply collective critical thinking to the generative outputs of LLMs.

The answer to the trust problem is in the next generation of AI—the Third Wave. The Third Wave of AI proposes a solution for explanation and context. We have illustrated the role of explanation in building trust, but how does context play a role?

Context is fundamental to understanding meaning. A word's meaning is highly dependent on its context. Transformer models are so named because they transform an input string of information into an output string. They learn context by tracking relationships between sequence components for large data sets. The combinatorics of this problem for large data sets leads to long, highly energy-consumptive training. The overall context window is 'all accessible information.' The context window for Large Language Model training is ever-increasing. This leads to the conundrum of insufficient data, causing model collapse. Open context then leads to energy issues, training time issues, and ultimate issues with model collapse. It is not a sustainable approach.

The path forward is to build on the first two waves of AI by embracing and extending what we learned. In the first wave, we studied human cognition in context and with explanation as a trust-building principle. In the second wave, we learned about the power of neural networks, statistical learning methods, and learning patterns of intelligence from data. We are poised for the third wave.

A framework for the third wave of AI

From the previous discussion, it is clear that we will reach the end of the road with statistical learning alone. As Judea Pearl, one of AI's most accomplished fathers, wisely stated, "You are smarter than your data. Data do not understand causes and effects; humans do." To achieve sustainable value from AI, we must include human creativity. Corporations, communities, and governing organizations are more than the data they leave behind. How do we tap into human creativity?

Adaptive knowledge acquisition of humans and AI agents working together requires thinking differently about AI methods. We can no longer use the 'batch' method of extensive training followed by deployment. We must learn as the brain learns: adaptively and continuously exploring cause-and-effect relationships. This adaptive learning method operates as a 'guide,' a 'facilitator,' and a 'meta-agent,' overseeing the collaborative work of humans and AI agents.

A new generation of AI builds on past successes in generative AI by creating a multi-agent architecture with a 'meta agent' that guides the collective intelligence of humans and AI Agents working together. It is based on two principles deeply rooted in physics and the natural sciences. One is the principle of emergent ordering from local interactions, and the other is the principle of least action — nature constantly seeks an efficient path forward. In addition, as in the natural sciences, it seeks to learn generative causal models rather than generating from statistical correlations.

We will now show how it is possible to ground a new generation of AI in nature. The innate intelligence of natural systems is the root of why AI works today.

The physics and biology of intelligence

Cooperative phenomena

When I first studied the physics of magnetic thin films in the late sixties, little did I realize that I was looking at something that had far deeper significance than the alignment of electron spins to create macroscopic magnetic effects. In 1974, two years after I finished my Ph.D., William Little published a fundamental paper on the link between magnetic systems and human memory.

> "The existence of such states of persistent order is directly analogous to the existence of long-range order in an Ising spin system, while the transition to the

state of persistent order is analogous to the transition to the ordered phase of the spin system."[4]

Eight years later, Hopfield published the paper, which led to his receiving the Nobel Prize in Physics this year. His paper built on Little's work to create the Hopfield model of the artificial neural network.[5]

Electrons have a property called spin, which leads to measurable macroscopic behavior from how spins interact. The Exchange Energy governs the behavior of electrons at the microscopic level as follows:

$$E = - J_{ij} \sum_{ij} s_i s_j \qquad (1)$$

Energy is minimized when spins are aligned, creating macroscopic measurable properties. We refer to the energy as 'free energy' - energy available for work. When the system is perturbed, it seeks equilibrium; in the case of a pure magnet, the spins align, and the minimum energy state results in a magnetic field. Equation (1) is fundamental to physics and underpins the principles of quantum information, leading to quantum computers.

In my work on creating a generative model to predict the behavior of magnetic thin films, I introduced other energy terms to represent the complexities of a multi-crystalline thin film. These terms included consideration of other forces that competed with the attractive force of alignment.

The Hopfield Model of the neuron mirrors each neuron as follows:

4 Little, W. A. (1974). "The existence of persistent states in the brain." Mathematical Biosciences, 19(1-2), 101-120.

5 Hopfield, J. J. (1982). "Neural networks and physical systems with emergent collective computational abilities." Proceedings of the National Academy of Sciences, 79(8), 2554-2558.

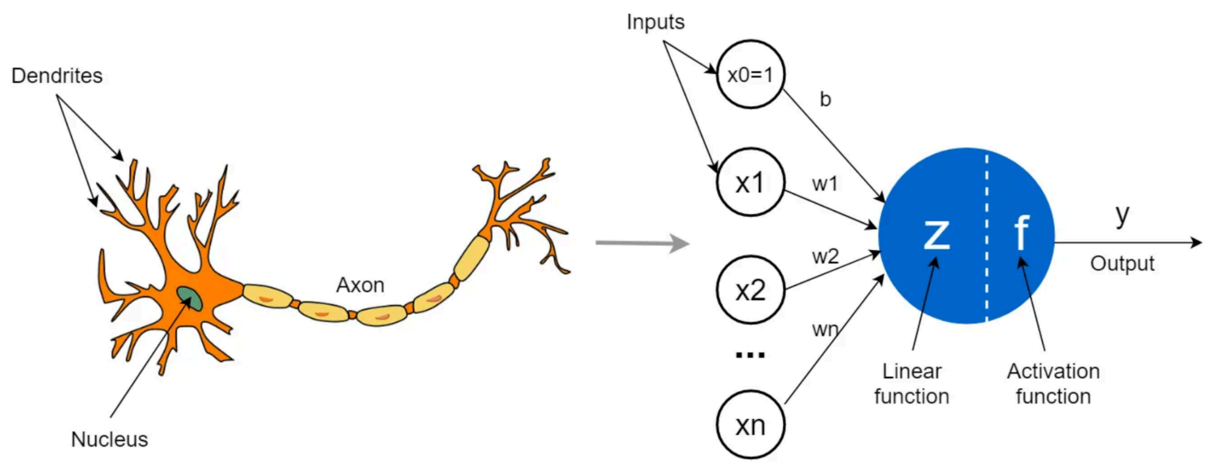


Figure 2[6]

The alignment of spins is equivalent to the alignment of inputs to a neuron, resulting in the following equation.

$$E = -\tfrac{1}{2} \sum_{i,j} w_{ij} x_i x_j - \sum_i \theta_i x_i \tag{2}$$

In the early development of the physics of magnetism, it was recognized that the principle of cooperative phenomena may be more general than first thought.

Recently, a group of MIT researchers concluded that deep learning works well because of its link to physics fundamentals. Various articles over the past few years have pointed out that the root cause for high performance was the link to nature.[7]

The figure below shows how this principle, derived from natural physics, has underpinned AI's accomplishments so far.

[6]https://towardsdatascience.com/the-concept-of-artificial-neurons-perceptrons-in-neural-networks-fab22249cbfc
[7]Lin, H.W., Tegmark, M. & Rolnick, D. Why Does Deep and Cheap Learning Work So Well?. J Stat Phys 168, 1223–1247 (2017). https://doi.org/10.1007/s10955-017-1836-5

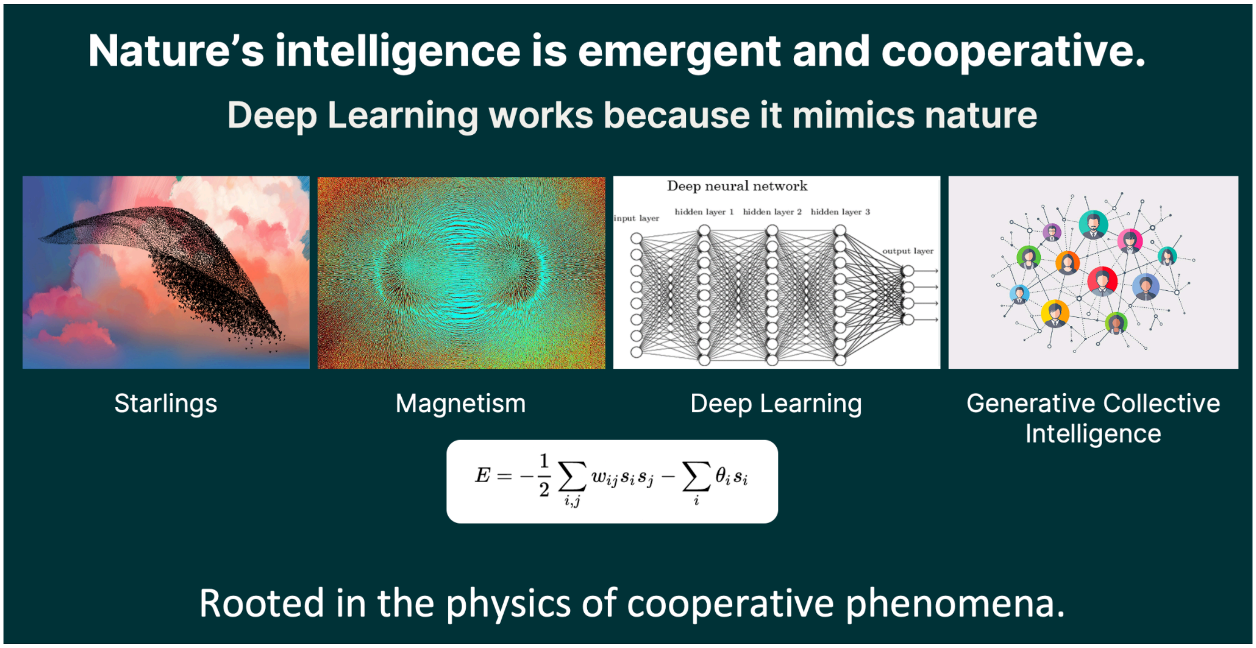


Figure 3

In the following sections, we will show how this applies to the third wave of AI, which will lead to a new form of generative AI: the generative collective intelligence of human and AI Agents working together.

> *Patterns of shared beliefs and knowledge follow the laws of cooperative phenomena!*

Applying cooperative phenomena to understanding patterns of belief and knowledge propagation has immense implications for the future of AI.

Least action - nature seeks an equilibrium (homeostasis)

The other physics principle related to the first of emergent order is the principle of least action—nature constantly seeks an efficient path forward. The principle of least action underpins the generation of how things work in physics.

$$\delta \int E = 0 \tag{3}$$

Where E is the energy of the system.

This principle leads to a new way of thinking about how the brain works to learn intelligently and navigate life.

The Free Energy Principle (FEP) and Active Inference

What does it take to stay alive? A pragmatic answer to that question leads to a foundation for AI. An intelligence has a generative model of what it needs to survive, processes sensory data, and makes choices that continually increase the probability of survival. The smart ones survive. In a humorous sense, the Darwin Awards are examples that question the intelligence of individual failure modes.[8]

The third wave of AI follows a model derived from the Free Energy Principle (FEP).[9] The FEP is a theoretical framework explaining how the brain generates its goals and desires based on sensory input. In its most straightforward formulation, the FEP and the accompanying process of Active Inference 'does science' with the environment to learn the best path forward to increase the likelihood of maintaining its existence. Active Inference is based on the idea that an agent should actively explore its environment to reduce the uncertainty about its internal state and the external world. Generative Collective intelligence is modeled as a collection of agents (humans initially) working collectively to plan a path forward with the highest probability of achieving a specific outcome.

Research in brain imaging led to the development of the Free Energy Principle. The theory is rooted in an AI model based on a fundamental principle in physics: all physical systems seek a place of rest (equilibrium). The principle is closely related to the principle of least action — nature seeks the most efficient way to get things done. Energy and information follow the same mathematical rules and living systems stay alive by reducing uncertainty in their choices. Minimizing free energy is the same as minimizing uncertainty in future choices. A model of how the FEP works in the brain is shown in Figure 2 below. In its simplest form, the brain compares sensory input to what it expects to observe, senses input, and compares expectations to observations. If sensory input is 'surprising,' the uncertainty of that surprise is called "free energy." It is mathematically identical to free energy in thermodynamics. Free energy is energy available to do work. Free Energy,

---

[8] https://darwinawards.com/

[9] Friston, K. The free-energy principle: a unified brain theory?. *Nat Rev Neurosci* 11, 127–138 (2010). https://doi.org/10.1038/nrn2787

viewed as uncertainty, implies that work must be done to resolve it. Active Inference is a process that seeks to reduce uncertainty by taking action.

> *At this point, one can see how this approach to AI is critical to organizations as learning communities seeking to forge a path forward that leads to agreed-upon goals and objectives. We are forging a path forward in making decisions about our future, whether for corporate or public good.*

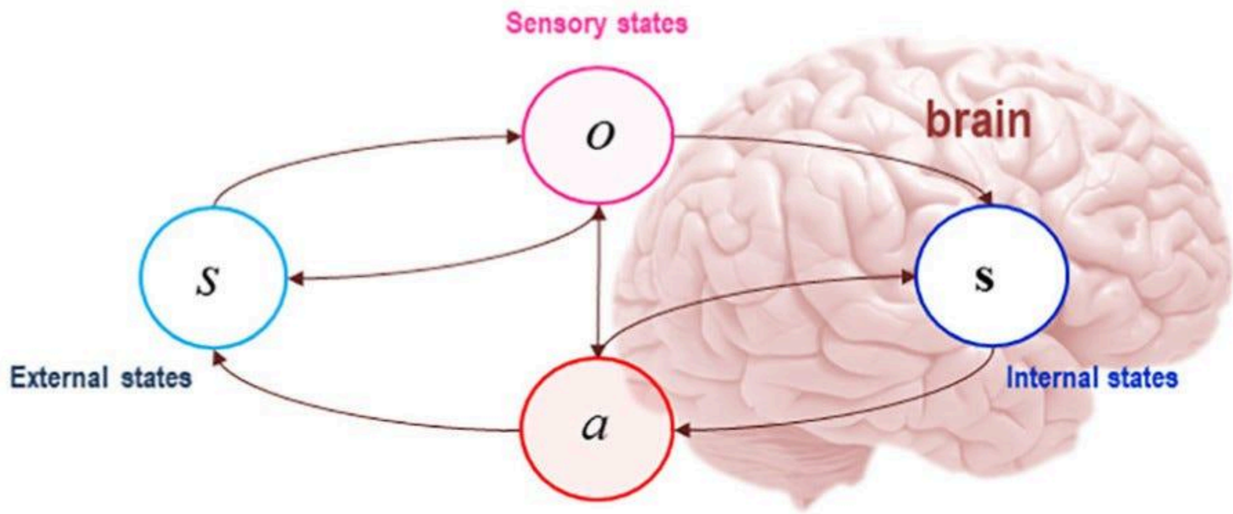


Figure 4

Expectations that drive perception (figuring out what we see is what we expected to see) are generated from our knowledge of the world — our beliefs about what we expect to see. This process is generative. Current AI is based on statistical predictions from the data products of our past. This new AI is a generative model of how humans (or other living systems) learn from their interactions with the environment — with the flow of events that define our life experiences. Fundamentally, it is an unfolding model of how we live in a complex world — how we learn and gain experience through a generative process of testing our beliefs (hypotheses) with the realities of life. A principal difference is model generation based on cause and effect vs. plausible statistical learnings from data.

The Science of Collective Intelligence

The theory of collective intelligence refers to the idea that groups of individuals can collaborate to achieve a level of intelligence and problem-solving capability that surpasses the abilities of any group member. Interpreted broadly, collective intelligence underpins all human achievements. Even the most celebrated human intelligence derives its achievements from the works of those gone before and of colleagues. We model the learning of collective intelligence as a Bayesian learning process.[10]

The essence of collective intelligence is a collection of intelligences will be smarter than any individual because of their cognitive diversity - they look at a problem from a variety of angles. In such a system, errors in individual judgment are offset by cognitive diversity. This is captured in the Diversity Prediction Theorem of Scott Page:

$$(c-\theta)^2 = 1/n\sum_{i=1}^{n}(s_i - \theta)^2 - 1/n\sum_{i=1}^{n}(s_i - c)^2 \tag{4}$$

Team collective error = Average individual error - Prediction Diversity

The AI model we will discuss in the following sections was initially developed in a specific context (early-stage investing) with a specific problem in mind (predicting the survival probability of a startup) - Can a group of cognitively diverse experts demonstrate the essence of this theorem? The answer is yes and is detailed in a paper published in 2018.[1] In creating the model, we evolved the concept of collective intelligence to include collective reasoning (cause and effect deliberation). In addition, we used a learning algorithm approach that mapped into the FEP and Active Inference.

Learning emergent alignment from collaborative deliberation

The problem of learning shared priorities is a known hard problem. Generally popularized as Brook's law, it states that learning shared beliefs about an outcome (shared cause and effect linkages) is highly complex. The mathematical formula for the lines of communication that must be managed is:

$$\text{Lines of communication} = n*(n-1)/2 \tag{5}$$

Learning aligned cause-and-effect relationships from 30 people requires managing 435 lines of communication. This is made more complicated by allowing every participant to generate new ideas based on what they have learned in the brainstorming/deliberation process. For this reason, Brook's law became known as the 'Mythical Man Month'. You cannot make a project go faster by adding more people because communication breaks down. In a later section, we will show that this problem is solvable by a probabilistic solution based on a solid mathematical foundation that is part of the new AI architecture. We will show that we can dynamically learn the emergent reasoning of deliberating intelligences in the new framework.

---

[10] Krafft, P. M., Zheng, J., Pan, W., Della Penna, N., Altshuler, Y., Shmueli, E., Tenenbaum, J. B., & Pentland, A. (2016). Human collective intelligence as distributed Bayesian inference. *ArXiv*. /abs/1608.01987

Managing context with a multi-criteria decision model

As mentioned, Large Language Models extend their intelligence by creating an extremely large context window, leaving context learning to highly costly and time-consuming statistical pattern recognition methods.

In practical terms, we acquire causal knowledge in specific contexts, solving specific problems. Tacit knowledge is a term used to describe what we sometimes call ‘intuition’ or ‘genius.’ Tacit is the intricate web of unspoken, intuitive, and experiential insights that individuals accumulate over time. Unlike explicit knowledge that can be easily codified and transferred through formal means, tacit knowledge is deeply personal and often challenging to articulate or document. Tacit knowledge gives rise to inspirational breakthroughs.

Tacit knowledge is most likely to be revealed through specific contextual deliberations. Ask an investor, “What is it about the team that you believe will make it successful?” Then, they will tap into a depth of tacit knowledge related to their experience.

A Multi-Agent Architecture for Third-Generation AI

We now have the framework for defining an AI architecture that solves the problems of the prior AI waves. The first wave is strong in explanatory power, managing context, and developing models of cognition. The first wave's failure points were the need for human development and maintenance. The first wave did not provide a means of self-learning and adaptation based on sensory data. The second wave is strong on statistical learning, deep learning, and transformer models, all based on learning from data. Handcrafting shifted from writing code to training data management and maintenance. In the case of transformer models, ‘aligning’ generative outputs to desired results still requires significant manual intervention. The third wave we propose here solves many of these problems through a multi-agent architecture with a special supervisory agent.

The approach proposed here is to use AI to learn collective knowledge models of human and (generative) AI Agents working together. The supervisory Meta-Agent learns causal models that retain provenance, explainability, and context. This is a critical component of bringing the scientific method to the evolution of AI.

The Agent Interface Environment (AIE) Architecture - Human and AI Agent Collaboration at Scale

The high level structure of the AIE architecture is shown in Figure 5. Human participants collaborate with each other and with agents vis the interface communication manager operated by a Meta-Agent. The Meta-Agent default mode is to hide the identity of agents forcing a focus on the intrinsic value of the knowledge shared rather than the identity of the generator of knowledge. This mimics the peer review process of scientific publication review.

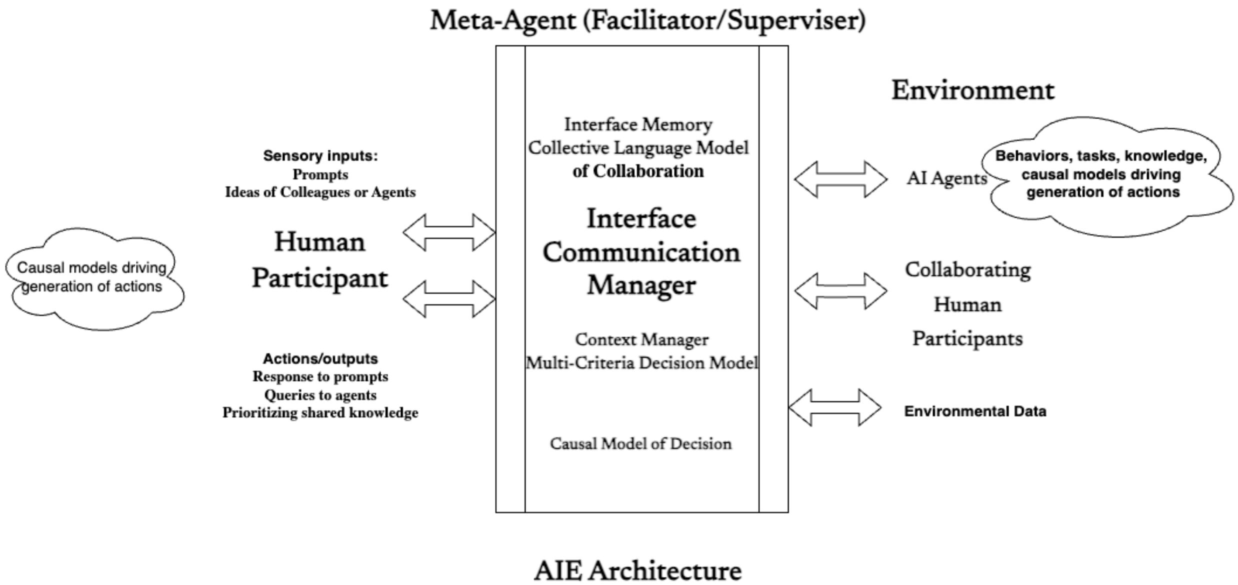


FIgure 5

The Meta-Agent orchestrates an interactive, collaborative process guided by a prompt script based on a multi-criteria decision model. The human participant on the left responds to a prompt using a Likert scale to estimate a criteria score. The participant is requested to give their (causal) reasoning for their estimate; they are then given a unique sample that includes comments from other human participants and AI Agents. It is at this point that the Meta-Agent is prompting to see if ideas from others instigate the synthesis of new ideas or approaches.

> ***In this step, we open the door to imaginative co-creation of human and AI agents working together.***

The goal is to learn the causal model of contributing agents. The figure shows a Human Participant on the left to illustrate the fact that any agent experiences the larger environment through the interface. The other participants, human or AI, are part of the environment from their point of view. The learning process is a partially observable Markov decision process (POMDP) using Bayesian mechanics as a learning guide.

The system builds a network of relationships through collaboration. In Figure 6 below, the outer nodes are participants. The links are relationships of aligned knowledge and priorities. The Meta-Agent builds a language model of the deliberation process. This model differs from current language models in that it retains the provenance of the knowledge including source (author's voice is retained) and context. it generates additional labeling of based on its learning from the deliberation process. Specifically it labels each comment/proposition with is estimated probability of relevance resulting from the peer review process. It also retains the network relationships of how each participating agent prioritizes its alignment with other agents.

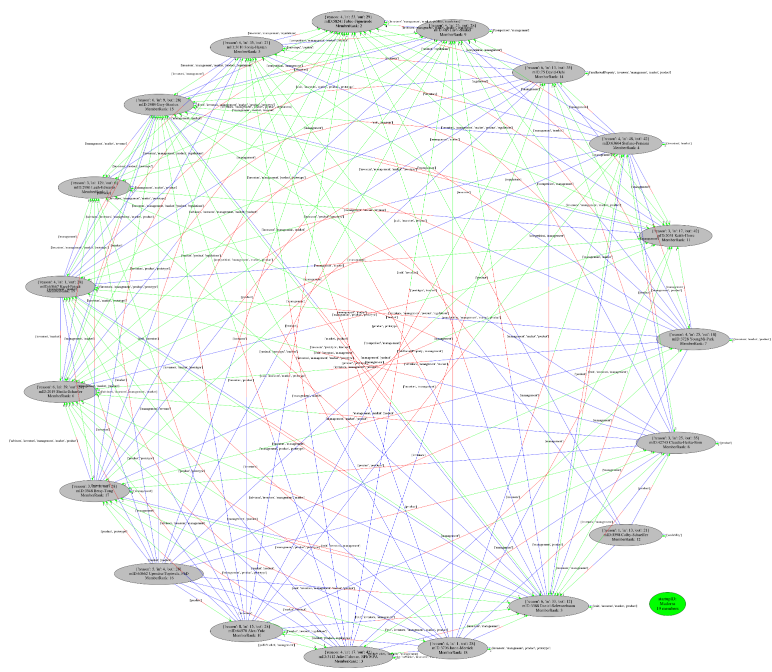


Figure 6

The system constructs a persistent causal model of the decision-making process from this network of relationships. In addition, it can create a private generative language model of the conversational deliberation process.

The AIE architecture extends learning beyond learning from explicit knowledge as articulated in text. Much of what drives human expertise is tacit knowledge that may only be observed through actions on the environment. Full exploration of the potential is beyond the scope of this paper.

Multi-criteria Model for the Investment Use Case

In this use case, we introduced the concept of collective reasoning to extract the tacit knowledge of investors when making a judgment or prediction for investing in an early-stage startup raising its first significant seed round (~ $1 million). The engagement process masked the identity of the collaborators during collective reasoning. Participants could see samples of other participants' reasoning but could not see their scores. The process was driven by a four-feature decision model that the system used to manage the process dynamically. If the predictive score exceeded a threshold, we would invest $50k to $100k (the decision was reviewed by an investment committee to monitor for error and suggest improvements to the process, but the decision was purely based on the score).

The results were published in a Neural Information Processing Society paper referenced above[1]. Results are shown in Figure 7.

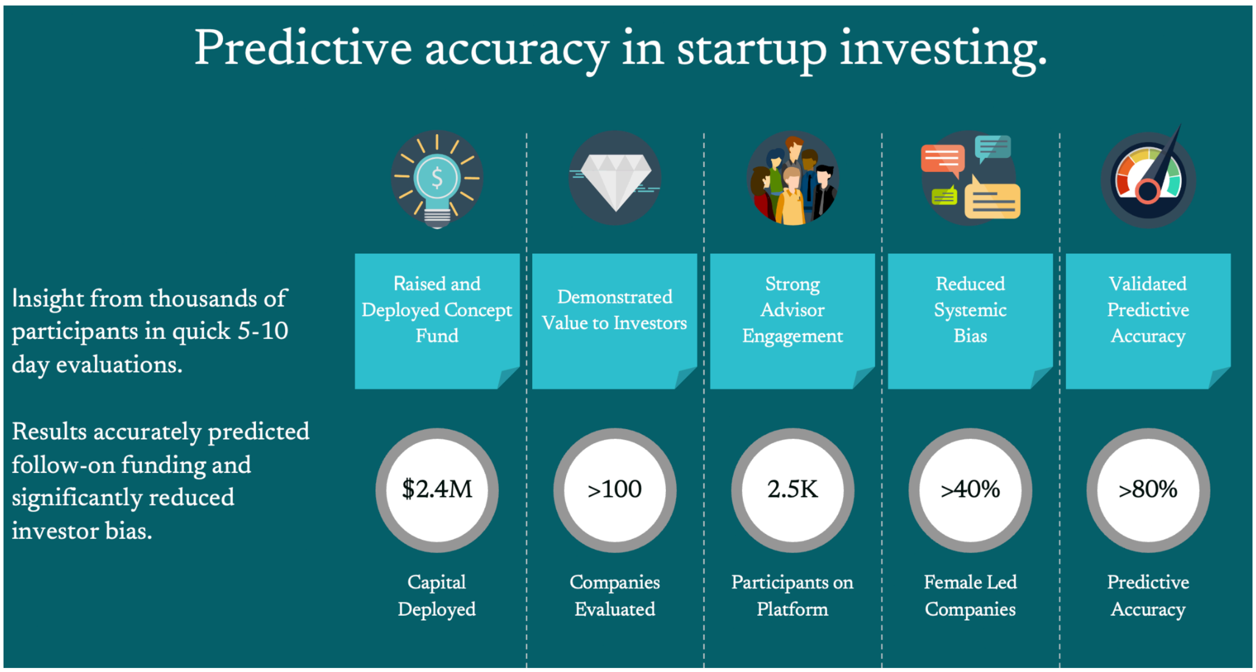


Figure 7

The system was shown to be >80% accurate in predicting the survivability of a startup as measured by the probability of raising follow on funding.

For each investment use case a language model of the deliberation was generated along with a causal model of the decision process. In the figure below the probability of survival generated from the deliberation process was 74%. The model parameters used by the Meta-Agent in guiding the collaboration were Business Opportuity, Team, Network Effects of Investors and Advisors and Investment Conviction. The green nodes denote clustered conversation topics learned from the deliberation. THe outer ring represents the scoring activities of the participants. The size of the node is correlated with with the degree to which the cause (evidence used) for the score is peer reviewed as relevant to the group. Scores with a high degree of peer-reviewed relevance are given greater weight.

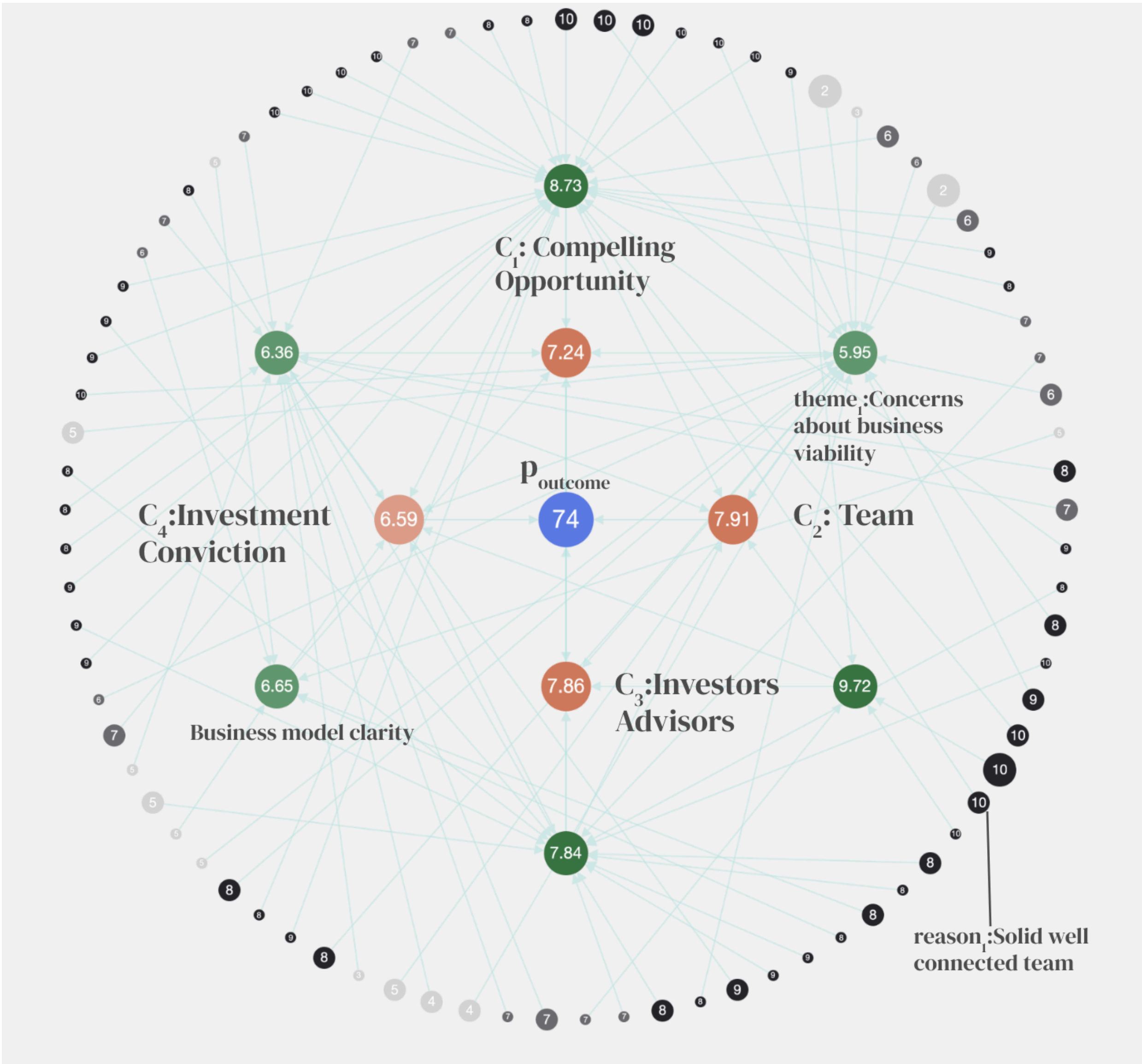


FIgure 8

The causal model is represented as a Bayesian Belief Network. It is a persistent representation of the decision process. It is available for simulation, execution (e.g. applicable to similar data streams), and generalization.

Creating Generative Collective Intelligence Models

We first look at learning shared preferences. The holy grail for preference learning is the A/B test or the law of comparative judgment. The law of comparative judgment underpins everything from eye tests to economic models predicting future choice. The meta-agent uses probabilistic adaptive learning methods to do this at scale. The figure is taken from a case where 100 people shared 3000 ideas in response to a prompt. An example output is shown below. Without a means of learning the shared priorities, all 3000 ideas are of equal importance — there is no clear signal forward.

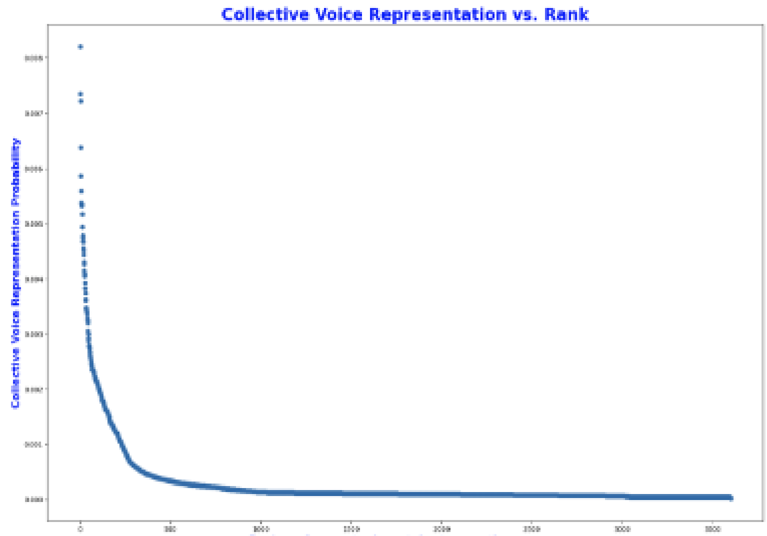


Figure 9

The system rapidly learns the top items of shared importance or relevance to the topic. Shared beliefs about an outcome build trust — even if the idea came from generative AI. Group alignment on ‘what is true’ or ‘evidence’ is fundamental to the scientific knowledge discovery method. The system discovers ideas that may have been generated due to the collaborative conversation.

> ***The combination of human creativity working with generative AI can foster an epiphany- a perception of a new path forward, a new way of seeing things, a way to break out on a new innovative solution.***

Various reporting tools are available to analyze all inputs. The system delivers the ability to learn top shared priorities and drive focused collaboration. It drives a large-scale deliberative conversation focused on shared goals and priorities, building trust as in the manual case. The output from one conversation drives the next. In addition, the system provides topic-level analysis and a generative query interface for the results.

How do we build trust in AI models?

Building trust in AI models involves several key strategies as outlined in the context: 1. **Scrutinizing Results for Errors**: Thorough examination of AI outputs is essential to ensure reliability and guide future directions. 2. **Balancing Scrutiny and Trust**: A synergy between AI and human creativity requires a balance, ensuring AI amplifies human intelligence without compromising data reliability. 3. **Explainable AI**: Making AI decisions more transparent helps assess data validity, which is crucial for building trust. 4. **Collaborative Improvement**: Working with talented individuals to continuously refine AI models enhances their accuracy and reliability. 5. **Data Validity and Provenance**: Ensuring the validity of training data and verifying data provenance are critical to maintaining trust. 6. **Addressing Hallucinations**: Minimizing hallucinations in AI outputs is vital for improving reliability. 7. **Synthesis Capability**: Trust in AI explanations is bolstered by their ability to synthesize vast information, enhancing collective decision-making. 8. **Ongoing Advancements**: Continuous improvements in guiding AI technologies reassure users about their reliability and future potential. By focusing on these strategies, trust in AI models can be effectively built and maintained.

View citations

This data was generated from a collaboration with mostly AI agents, who were given different personas. A few humans participate in the conversation. Each narrative summary is supported by a 'collective voice' analysis, which shows the prioritized importance of each comment that led to this summary.

25 In my experience, scrutinizing results for errors is essential to guide the future direction of AI technologies. A thorough examination ensures that AI models like LLMs deliver reliable insights, which is crucial for market formation and innovation.

Rating: 6/10

Reply to author

The system tracks who said what, when, and how relevant or important the statement was through the peer-reviewed process described previously. The 25 is the probability estimate that this is relevant to the entire group. Any participant can also query the source (whether an AI agent or a human) and drill down on evidence.

Thinking Together with Generative Collective Intelligence

For organizations to build trust in AI, we must add the word ‘collective’ to Generative Intelligence or Generative AI. Generative Collective Intelligence takes inspiration from the intersection of design thinking and collective intelligence. Collective intelligence is the group’s shared intelligence that emerges when people work together, often with the help of technology, to solve complex problems. “Collective intelligence is believed to underlie the remarkable success of human society. “[1] The application of AI techniques to make collective intelligence work at scale requires an AI capability of collaborative adaptive learning. It is an entirely different approach to how we think about AI — a fundamental shift from learning from the data trails of human intelligence to becoming an integral part of the collaborative prediction, planning, and problem-solving of humans and AI Agents.

Generative Collective Intelligence Applied to Enterprise Decision-Making

To illustrate the practical application of Generative Collective Intelligence, we consider a use case of an established company making a ‘build or buy’ decision. A well-established company in healthcare information systems is challenged by a significant shift in healthcare to the distributed mobile acquisition of patient information. The company is considering acquiring or creating a relationship with a fast-growing startup or increasing investment in an internal project. There is a wide variety of stakeholders with diverse expertise. (In this case, AI Agents are not used.) They realize this is far more than a financial decision if they want to make it work. They want to engage a larger group of stakeholders to gain intelligence on the best path forward. They have created a multi-criteria decision model and have engaged the diverse teams in a collaborative effort that involves deep discussion and deliberation, all orchestrated by the Meta-Agent. The Meta-Agent is given a script to guide the collaboration process. The Meta-Agent learns the shared preferences of the group through a process of interacting with each individual, learning how their priorities align with those of their colleagues.

The result is a ‘collective voice’ of the group that represents their collective response to every area under investigation. Hundreds of ideas are shared and discussed. The system learns what is of shared importance to the group. For example, one of the areas of consideration is risk:

Integration Risks

Filters (0)

How well suited is the company to to integrate into the mainstream business (overcoming cultural and operational risk)

Collective Voice

Collective Voice shows us the ideas that best represent the thinking of the group. It looks at all the ideas that were entered for a question or collaboration and shows you what represents the group's voice based on how important people ranked those ideas.

1. At this stage in our diligence the legal risks of their current business have been assessed and addressed, yet their 'cowboy' style leaves me with lingering concerns about their potential for internal disruption. They are either unfamiliar with or lack appreciation of risk managment practices that are essential to our role in the healthcare system. Suggest we consider alternatives to fully assimilating them into the organization

2. The engineering organization is geographically dispersed, with at least two clusters of developers in eastern europe. This could pose significant new challenges to us in coordinating future development and keeping people onboard as they shift from a highly independent culture to a more integrated, process driven team on a much larger scale

3. This team is sharp, I respect them and what they've built. That being said I'm having difficulty imagining them enjoying the indoctrination into the corporate processes. Suggest they stay as a stand alone subsidiary for eveyrone's best interests

More importantly, a generative interface is available that captures the group's collective intelligence, which can then be broadly queried.

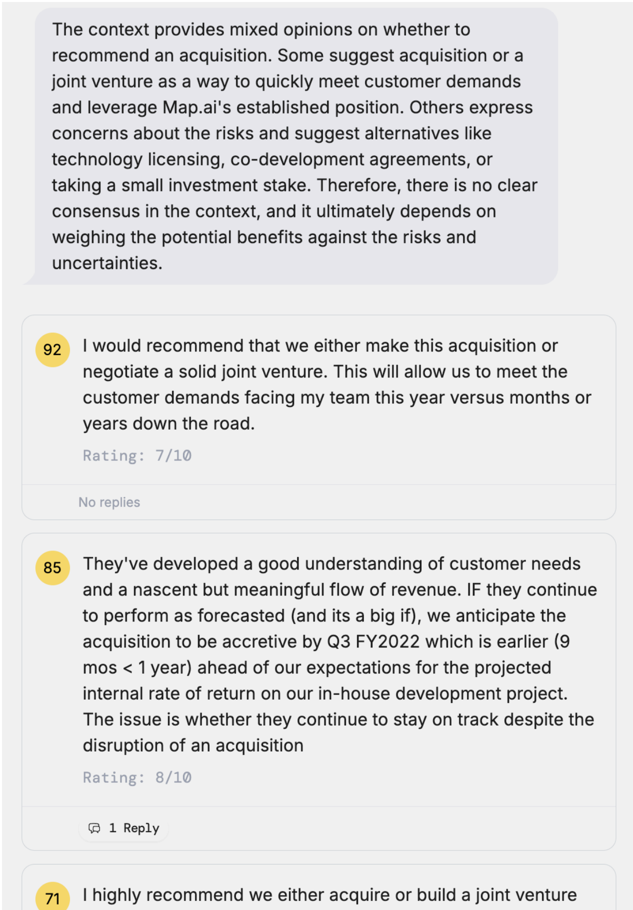
Do you recommend acquisition?

The context provides mixed opinions on whether to recommend an acquisition. Some suggest acquisition or a joint venture as a way to quickly meet customer demands and leverage Map.ai's established position. Others express concerns about the risks and suggest alternatives like technology licensing, co-development agreements, or taking a small investment stake. Therefore, there is no clear consensus in the context, and it ultimately depends on weighing the potential benefits against the risks and uncertainties.

CS

View citations

The original voice of each participant is preserved, so summaries are backed up with citations.



In addition to the language interface, the Meta-Agent builds a causal model of the entire decision process that creates an 'audit trail' on decisions, which is useful in improving decision accuracy. For certain categories, knowledge models derived from repeat use cases can be used to automate decision processes.

The path forward is Generative Collective Intelligence based on first principles.

Taping into the power of emergent correlations in data has created great advances. Recently, Hopfield and Hinton were awarded Nobel Prizes in physics for their significant work in neural networks and deep learning. Hinton and Bengio, founders of deep learning, have clarified that while deep learning is a substantial move in the right direction, it is not the complete answer to AI.

> "In terms of how much progress we've made in this work over the last two decades: I don't think we're anywhere close today to the level of intelligence of a

> 2-year-old child. But maybe we have algorithms that are equivalent to lower animals for perception."[3]

We are on a new road forward when we brace Generative Collective Intelligence because it *embraces the scientific discovery process.* It offers the ability to build trust and embrace our creative intelligence.

We have over-indexed on seeing learning statistical patterns from data as the future of AI. It is a false hope to believe that what we see today will lead to a positive future for AI. There is reason for caution and concern because we are not there yet. The answer is not to constrain or regulate but to see the bright future ahead by investing our energies in much richer models based on the first principles we can learn from physics and the natural sciences. As Richard Feynman states:

> "Although we humans cut nature up in different ways, and we have different courses in different departments, such compartmentalization is really artificial... **The imagination of nature is far, far greater than the imagination of man."**
> ~Richard Feynman

Summary (TL: DR)

Trust in AI for the future will only happen with an ability to apply collective critical thinking to results. Third-generation AI addresses issues of trust explanatory power and opens the door to co-creation between humans and AI agents, building a process of collective reasoning and trust-building rooted in the scientific knowledge discovery method. In addition, multi-agent architectures that adaptively learn are sustainable and can be designed to be more energy efficient than large language models are today.