Journal of Creativity 34 (2024) 100077

Contents lists available at ScienceDirect

Journal of Creativity

journal homepage: www.elsevier.com/locate/yjoc

AI for Technoscientific Discovery: A Human-Inspired Architecture

J.Y. Tsao [c,*], R.G. Abbott [b], D.C. Crowder [b], S. Desai [c], R.P.M. Dingreville [c], J.E. Fowler [c], A. Garland [a], P.P. Iyer [c], J. Murdock [b], S.T. Steinmetz [b], K.A. Yarritu [a], C.M. Johnson [a], D.J. Stracuzzi [b]

[a] Applied Information Sciences Center
[b] Center for Computing Research
[c] Material, Physical and Chemical Sciences Center, Sandia National Laboratories, Albuquerque, NM

ARTICLE INFO

Keywords:
Artificial intelligence
creativity
extended neurosymbolic knowledge
technoscientific method

ABSTRACT

We present a high-level architecture for how artificial intelligences might advance and accumulate scientific and technological knowledge, inspired by emerging perspectives on how human intelligences advance and accumulate such knowledge. Agents advance knowledge by exercising a technoscientific method—an interacting combination of scientific and engineering methods. The technoscientific method maximizes a quantity we call "useful learning" via more-creative implausible utility (including the "aha!" moments of discovery), as well as via less-creative plausible utility. Society accumulates the knowledge advanced by agents so that other agents can incorporate and build on to make further advances. The proposed architecture is challenging but potentially complete: its execution might in principle enable artificial intelligences to advance and accumulate an equivalent of the full range of human scientific and technological knowledge.

Introduction

Recent advances in artificial intelligence (AI) have been nothing short of extraordinary (Bengio et al., 2021). Inspired by those advances, interest has been growing in artificial intelligence applied to one of the most distinctive features of human intelligence: the ability to advance and accumulate scientific and technological knowledge (LeCun, 2022; Stevens et al., 2019; The Economist, 2023; Wang, Fu, et al., 2023)—what we will refer to as technoscientific knowledge (Latour, 1987).

In this paper, we consider this potential for AI to execute, basically, the equivalent of human scientific and engineering research and development, including, importantly, creative discovery. The result is a high-level architecture by which artificial intelligence might mimic human intelligence in advancing technoscience. That said, in borrowing from the human, we by no means wish to imply that the artificial *must* mimic the human. But, just as current AI architectures take inspiration from human intelligence—from the abstracted neuron as the foundational element for deep neural networks (James et al., 2017) to cognitive architectures as infrastructures for intelligent agent systems (Newell, 1994)—AI that advances and accumulates technoscientific knowledge might fruitfully take inspiration from how human intelligence does the same.

A high-level view of the architecture is sketched in Fig. 1. In the upper half of the figure, an agent interacts with, and advances technoscientific knowledge about, the world. Agents are rewarded for their knowledge advances by both extrinsic (Merton, 1973) and intrinsic (Hennessey et al., 2015) rewards. In the bottom half of the figure, the agent shares technoscientific knowledge with society. Agents transfer knowledge to society so that knowledge can be accumulated and passed from generation to generation even as individual agents are born and die (Henrich, 2015). Society transfers knowledge back so that agents can build on society's accumulated knowledge as they interact with, and advance technoscientific knowledge about, the world (Deutsch, 2011). This architecture represents a kind of decentralized (Hayek, 1945) but collective (Watson & Levin, 2023; Wolpert & Tumer, 1999) intelligence that, at a high level, is descriptive of human intelligence as it advances and accumulates technoscientific knowledge.

The remainder of this paper outlines this architecture. First, we quantify the rewards that agents might receive as they advance technoscience—what we call "useful learning." Second, we discuss the structure of technoscientific knowledge, and the various knowledge representations needed as knowledge is advanced and accumulated. Third, we describe the mechanisms by which technoscientific knowledge is advanced. Taken together, these three sections describe a

* Corresponding author
E-mail address: jytsao@sandia.gov (J.Y. Tsao).

https://doi.org/10.1016/j.yjoc.2024.100077
Received 8 November 2023; Received in revised form 27 January 2024; Accepted 28 January 2024
Available online 8 February 2024

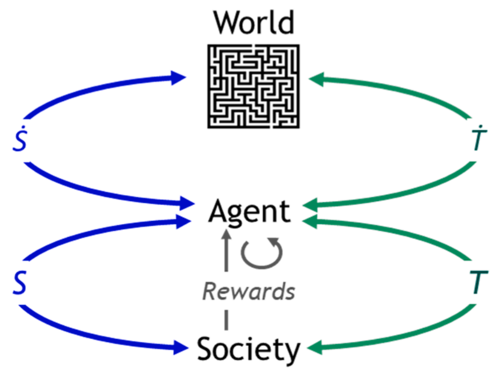

Fig. 1. High-level view of an architecture for how human agents advance and accumulate technoscientific knowledge. Facing upwards in the figure, individual agents interact with, and advance technoscientific knowledge about, the world. To convey the sense that technoscientific knowledge advance necessitates or entails change, we denote such scientific and technological advance using the symbols $\dot{S}$ and $\dot{T}$, respectively. Facing downwards in the figure, individual agents interact with (learning from and teaching) society, which accumulates technoscientific knowledge that is handed down from one generation of agents to the next. To convey the sense that technoscientific knowledge is accumulating in repositories, we denote these accumulations using the symbols S and T.

complete, if high-level, architecture for the advance and accumulation of technoscientific knowledge, including advances that are creative. Fourth and finally, we discuss examples in which artificial intelligence is already being used to execute subsets of this architecture.

Our architecture weaves together perspectives from a wide range of disciplines, including: the history, philosophy and social science of science and technology; cognitive and social psychology; human creativity; and collective and embodied intelligence. Thus, we recognize that the overarching architecture may seem foreign to any one of these disciplines, not to mention to the AI community itself. Nevertheless, our hope is that the architecture will provide a useful overarching perspective on an exciting emerging area of modern AI: AI for technoscientific discovery.

Knowledge Advance as Useful Learning

In this section, we discuss how we might quantify knowledge advance—what we will call useful learning. We draw heavily on recent treatments of knowledge advance (Simonton, 2018; Tsao et al., 2019) in which useful learning is expressed as the product of utility and learning, $u{\cdot}l$. Useful learning can only be high if both utility and learning are high—if either is low, than useful learning is low.

Useful Learning: Implausible and Plausible Utility

To concretize this concept of useful learning, consider Fig. 2(a). Suppose we have a potential new nugget of knowledge. Suppose we make guesses as to the utility of that potential nugget of knowledge. Along the left axis, in red, is plotted a guess made *before* the potential nugget of knowledge is tested. Along the right axis, in blue, is plotted a guess made *after* the potential nugget of knowledge has been tested. In this particular scenario, the prior probability distribution is broad and at the lower end of the utility scale, while the posterior probability distribution is narrower and at the higher end of the utility scale. This means that the testing has resulted in learning—the potential new nugget of knowledge is more useful than was originally thought, and there is less uncertainty about that usefulness.

To measure how much has been learned, we use the Kullback-Leibler (KL) divergence (D_{KL})—a standard measure of the difference between the prior and posterior probability distributions. How much one has learned depends on two key characteristics of the two distributions, both defined in Fig. 2(a): Δb, blindness reduction, is the normalized change in the widths of the distributions; and s, surprise, is the normalized change in the mean values of the two distributions. The result is the following approximate equation for useful learning (Tsao et al., 2019):

$$u{\cdot}l \;=\; u{\cdot}D_{KL}\left(p_{prior}, p_{post}\right) \;=\; u{\cdot}D_{KL}(\mathrm{s}, \Delta b) \;\approx\; u{\cdot}s^2 \;+\; u{\cdot}ln(1+\Delta b)$$

where, for simplicity, we have replaced the $\bar{u}_{post}$ defined in Fig. 2(a) with u.

The first term in the final expression, $u{\cdot}s^2$, contains surprise and, because surprise is what we might call implausibility, we can think of this term as implausible utility. This term is exemplified by the scenario drawn in Fig. 2(a), in which the potential nugget of knowledge was thought not to be useful (its utility was disbelieved), but it turned out to be useful (that disbelief was disconfirmed). The second term in the final expression, $u{\cdot}\ln(1+\Delta b)$, contains blindness reduction but not surprise. Because absence of surprise is what we might call plausibility, we can think of this term as plausible utility.

Both terms, implausible and plausible utility, are important, but they play different roles in the landscape of research and development. Implausible utility is more creative and more research-like (disruptive and more akin to “revolutionary” science and engineering), while

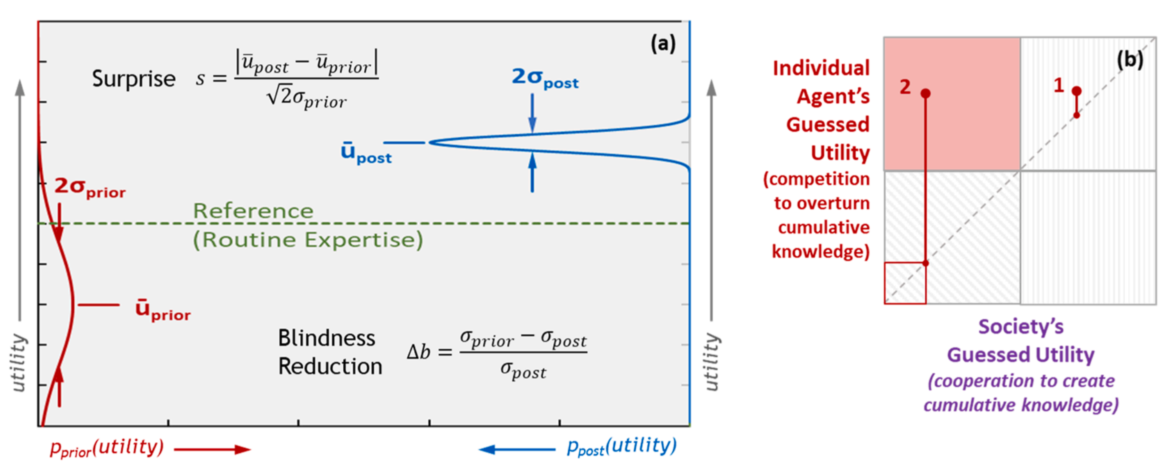

Fig. 2. (a) Probability distributions associated with guesses about the utility of a proposed nugget of knowledge prior (left, in red) and posterior (right, in blue) to testing. Two key characterstics of the two distributions, surprise (s) and blindness reduction (Δb) are defined in the figure. (b) An individual agent’s versus society’s guesses as to the utility of that proposed nugget of knowledge, before testing. These two guesses need not be the same and the more they differ the greater the potential for surprise to, and learning by, society.

plausible utility is less creative and more development-like (consolidative and more akin to "normal" science and engineering) (Dosi, 1982; Kuhn, 1962). They also have different bibliometric signatures (if published as traditional journal articles)—the former "obliterating" and the latter maintaining, citations to previous work (Funk & Owen-Smith, 2017; Leibel & Bornmann, 2023).

Useful Learning to Whom: Agent vs Society

Useful learning, we posit, is a potential measure of knowledge advance. But to whom is this useful learning in reference to: the agent or society? Ultimately, of course it is to society: societal knowledge is what is cumulative, what is passed from one generation to the next. Knowledge that is uniquely held in individual agents, not communicated to, or otherwise kept separate from, society, becomes lost when the agent ceases to function (or, if human, dies).

What do we mean by usefulness to society? At a high level, we view technoscientific knowledge as interconnected in a "seamless web" of knowledge (Anderson, 2001). Utility, then, can perhaps be measured as the degree to which a nugget of knowledge is embedded in society's web of knowledge—how useful that nugget of knowledge is to the rest of the web of knowledge. Maxwell's equations are useful because they connect to classical electromagnetic phenomena in fields ranging from astrophysics to ionic transport in cells. Integrated circuit chips are useful because they connect to computation in applications ranging from interplanetary missions to iPhones.

What do we mean by learning, especially that associated with creative discovery and implausibility to society? As discussed above, we mean by this *surprise* to society. Importantly, though, surprise to society does not necessarily imply surprise to the agent. To see why not, consider Fig. 2(b). The bottom axis depicts what society thinks the utility of a potential new nugget of knowledge might be; while the left axis depicts what the individual agent thinks the utility of a potential new nugget of knowledge might be. Two scenarios are depicted. In scenario 1, both agent and society assess utility as high and there is not much opportunity for surprise. Even if the individual agent is right that this potential nugget of knowledge will be useful, it will not be surprising to society because society already thinks it will be useful. In scenario 2, society assesses the utility as low, but the individual agent assesses the utility as high. This disagreement *is* opportunity for surprise to society. The agent is acting as a contrarian, disagreeing with society's assessment of utility for reasons that it knows, but society does not. It is when the agent is such an informed contrarian that there is a good chance the agent is correct and society is incorrect, and opportunity for surprise to society is maximized (Boden, 2005; Simonton, 2013; Thiel & Masters, 2014).

Technoscientific Knowledge

In the last section, we discussed knowledge advance as useful learning. In this section, we particularize knowledge advance and useful learning to *technoscientific* knowledge. We begin with a definition of technoscientific knowledge as the sum of scientific and technological knowledge, each with a similar underlying structure. Then we discuss how, because of the different roles agents and society play in the advance and accumulation of technoscientific knowledge, they must rely on different representations of that knowledge.

Structure of Technoscientific Knowledge

We begin with the structure of technoscientific knowledge. As discussed recently (Narayanamurti & Tsao, 2021), and as outlined in the top row of Fig. 3, this knowledge can be thought of as composed of two repositories: science and technology. The science and technology repositories in turn can each be thought of as composed of two bins. The two bins of the science repository, *S*, in blue at the top left of Fig. 3, are facts of human interest and explanations of those facts. The two bins of the technology repository, *T*, in green at the top right of Fig. 3, are human-desired functions and forms that fulfill those functions.

At a deeper level of structure, the bins themselves can be thought of

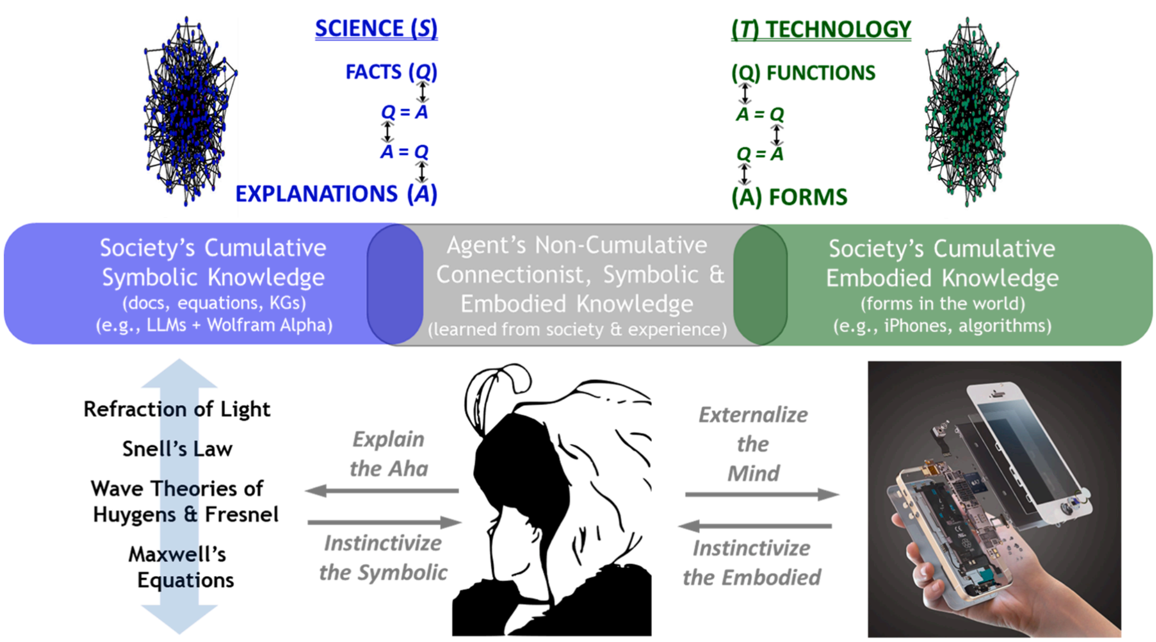

Fig. 3. A sketch of how technoscientific knowledge might be structured, represented, and translated between agents and society. The top row illustrates scientific knowledge (facts of agential interest and explanations of those facts) and technological knowledge (agent-desired functions and forms that fulfill those functions). The middle and bottom rows illustrate society's symbolic scientific knowledge at the left, society's embodied technological knowledge at the right, and an individual agent's non-cumulative connectionist, symbolic and embodied knowledge in the middle. Society's symbolic scientific and embodied technological knowledge are cumulative and can be passed on from one generation of agents to the next. An agent's connectionist knowledge must be translated into symbolic or embodied knowledge in order to be transferred to, and accumulated by, society.

as recursively nested into loosely modular hierarchies of question-answer pairs (Anderson, 1972; Arthur, 2009; Simon, 1962). As illustrated at the bottom left of Fig. 3, the refraction of light is an observed fact at one level, explained by Snell's law one level below, in turn explained at a deeper level by the wave theories of Huygens and Fresnel, and then at an even deeper level by Maxwell's equations. As illustrated at the bottom right of Fig. 3, an iPhone is a physical form that fulfills the human-desired function of portable computing and communications. But it also represents an integrated collection of functions that require sub-forms, like Gorilla Glass or integrated circuits, to fulfill; and these sub-forms represent functions that, in turn, require sub-sub forms to fulfill.

Representations of Technoscientific Knowledge

As discussed in the Introduction, agents and society each play important, but different, roles: agents advance, while society accumulates, technoscientific knowledge. On the one hand, the accumulation of knowledge requires that knowledge be easily shared between individual agents and society—it requires knowledge representations that are compact and manipulable. On the other hand, the advance of knowledge requires deep and intimate interactivity with the world—it requires knowledge representations that capture the complexities of the world.

As depicted in the middle row of Fig. 3, one possibility for how human agents and societies, and how artificial agents and their societies might, execute their different roles is to make use of three fundamentally different knowledge representations. At the left, in blue, society's cumulative scientific knowledge is depicted as largely symbolic, while at the right, in green, society's cumulative technological knowledge is depicted as largely of a type that might be called "embodied." At the center, in grey, agential scientific and technological knowledge is depicted as largely connectionist, with some overlap with the symbolic and the embodied. Two of these knowledge representations, the symbolic and connectionist, have been discussed in the context of neurosymbolic representations and Kahneman's "think fast" and "think slow" (Bengio et al., 2021; Booch et al., 2021; Kahneman, 2011; Kambhampati, 2021; Marcus, 2003). Here, we extend this to include the embodied, in green at the right. The symbolic and embodied representations allow society to accumulate knowledge, and the connectionist representation assists agents as they go beyond that accumulated knowledge. Building on the notion of the extended mind (Clark & Chalmers, 1998)—a mind that includes aspects of the external world—we might call the sum of these three knowledge representations *extended neurosymbolic knowledge*.

Consider, first, society's cumulative scientific knowledge, depicted as largely symbolic. This knowledge representation is formal, logical, reductionist, and explanatory (Hadjimichael, 2023; Polanyi, 1966). Most importantly, it is teachable from one agent to another and so is cumulative, outliving individual agents. The symbolic knowledge inherent in knowledge graphs and engines such as Wolfram Alpha (Davis & Aaronson, 2023) and in equations such as $F = ma$ can be taught by one agent to another. Symbolic knowledge serves as the repository of society's cumulative scientific knowledge, hence enables Isaac Newton's "If I have seen further it is by standing on the shoulders of Giants" (Newton, 1675).

Consider, second, society's cumulative technological knowledge, depicted as largely embodied in the actual forms and functions of technology (Clark & Chalmers, 1998; Tanaka, 2011; Valiant, 2013). Much like there is knowledge encoded in biological life—a bird's knowledge of how to fly is in part encoded in the structure of its muscles, bones, and nervous system—there is knowledge encoded in technology. A hammer—how it is shaped and constructed—encodes knowledge about nails and the human hands that drive them. In other words, technologies encode knowledge about the humans that are using them and the environments within which they are being used (Gibson, 1977). Most importantly, such knowledge is transferrable from one agent to another. The embodied knowledge inherent in a hammer is transferred when one agent physically gives the hammer to another agent. Embodied knowledge is what enables an agent with an iPhone to be more productive than an agent without an iPhone, whether or not the agent understands the scientific principles underlying iPhones.

Consider, third and finally, agential knowledge, which does not accumulate beyond the agent, but whose purpose is to enable agents to interact with the world and, as they do so, to advance knowledge. We view agential knowledge largely as connectionist, tacit knowledge (Polanyi, 1966), but with extensions into the symbolic and embodied. It is the detailed and intuitive knowledge that comes from direct interaction with the world and that enables the agent to navigate quickly and accurately in the world. It is the recognitional knowledge that enables the agent to instantly see either the young or old woman in the Gestalt sketch illustrated in Fig. 3, but not at the same time. The rich, connectionist details enable Einstein's: "Imagination is everything. It is the preview of life's coming attractions" (Viereck, 1929). Imagination may lead to hallucinations, but may also yield breakthrough insights and surprising new ways of thinking. Switching from seeing the young woman to seeing the old woman could be just the aha that solves some puzzle at hand.

Translating between Knowledge Representations

As just discussed, the three knowledge representations (symbolic, embodied, and connectionist) are necessary but for different purposes: the symbolic and the embodied facilitate society's accumulation of knowledge, and the connectionist facilitates agents' going beyond that accumulated knowledge. This means agents and society need to be able to transfer knowledge to and from each other, and knowledge needs to be translated between representations, as indicated by the arrows in the bottom row of Fig. 3.

Consider first translation between the connectionist and the symbolic. In one direction, translation from the connectionist to the symbolic might be thought of, for artificial intelligence, as "explainable AI" (Roscher et al., 2020), or, for human intelligences, as "explain the aha." It is not enough for a connectionist agent to have a new insight; the agent must represent the insight symbolically so that it can explain the insight to others. In the other direction, translation from the symbolic to the connectionist might be thought of as "instinctivize the symbolic." For artificial intelligence, a connectionist neural network might be trained on materials structure-property relations calculated using computationally expensive symbolic physics, and the patterns it has learned or "instinctivized" from those relations might then be used to calculate the properties of new structures using computationally inexpensive inference (Schleder et al., 2019). For human intelligences, an analogy might be: students learning some physical phenomenon via symbols and equations, initially being able to solve problem sets but without intuitive understanding, then having an aha moment in which suddenly they understand what the symbols and equations actually "mean." There is a difference between a symbolic, perhaps mathematical, *explanation* of a phenomenon versus a connectionist, intuitive *understanding* of it (De Regt, 2017; Krenn et al., 2022).

Consider, second, translation between the connectionist and the embodied. In one direction, translation from the connectionist to the embodied might be thought of as the actual embodiment of ideas into real-world forms which perform real-world functions—a kind of "externalization of the mind" (Clark & Chalmers, 1998). The connectionist does not *need* to be translated into the embodied. But, when it is, it can add to society's cumulative knowledge, and when it is not, it dies when the agent dies. In the other direction, translation from the embodied to the connectionist might be thought of as a kind of "instinctivizing of the embodied." By directly experiencing the technology, by actually riding the bicycle or hammering the nail, a kind of connectionist and intuitive knowledge is created in the agent (Johnson, 2015). Such "instinctivizing the embodied" can be extremely powerful.

Our use of synthetic sensory and actuation technologies, like prosthetic ears or hands, can become nearly as hard-wired into the human brain as our use of our biological ears and hands (Eagleman, 2011).

Technoscientific Knowledge Advance

In the last section, we discussed the structure and representations of technoscientific knowledge. Agents learn from society (knowledge is transferred from society to agents) and combine that knowledge with new experiences in the world to advance (create new) knowledge. In this section, we outline how agents advanced knowledge by finding scientific facts and explanations and/or by finding technological functions and forms. We borrow heavily from an emerging perspective in which agents find facts/explanations and functions/forms via a so-called technoscientific method (Narayanamurti & Tsao, 2021).

The Technoscientific Method

What is the technoscientific method? It is the interacting combination of a scientific method, $\dot{S}$, and an engineering method, $\dot{T}$, each composed of three mechanisms.

The scientific method is shown at the left of Fig. 4. Though there is no particular beginning or ending point to the scientific method, for the sake of starting somewhere we start with $\dot{S}_1$. This is the finding of facts, both new facts that go beyond existing theory as well as new facts intended to test emerging theory—the first type exemplified by Galileo's discovery of the moons of Jupiter and the second type being classic hypothesis testing. With facts in hand, the method proceeds to $\dot{S}_2$, the finding of explanations for those facts—this is classic theorizing or hypothesis generation, as in Einstein's explanation of the constancy of the speed of light by the theory of relativity. And then, with emerging explanations in hand, the method proceeds to $\dot{S}_3$, the generalizing of those explanations to predict possible new facts, triggering a hunt for those new facts—coming full circle back to the half of $\dot{S}_1$ that is hypothesis testing, as in Meitner's discovery that special relativity could explain energy release during nuclear fission.

The engineering method is at the right of Fig. 4, and is exactly analogous to the scientific method. The method starts with $\dot{T}_1$, the finding of human-desired functions. Practical utility may be behind some of the functions, as in Steve Jobs' idea of the human desirability of the functionality of the iPhone (Merchant, 2017). But curiosity and learning may also be behind some of the functions, as in Albert Michelson's desire to measure the reference-frame-dependence of the speed of light (Harré, 1981). Then, with functions in hand, the method proceeds to $\dot{T}_2$, the finding of forms that fulfill those functions, as in the Wright Brothers' invention of heavier-than-air flight. This brings us to $\dot{T}_3$, exapting, the co-opting of existing forms to fulfill functions they were not originally intended to fulfill—for example, Percy Spencer's co-opting of radar tubes for the invention of microwave cooking (Osepchuk, 1984). The term "exapting" (Andriani & Carignani, 2014) is borrowed from evolutionary biology, in which biological forms, like dinosaur feathers for thermoregulation, are co-opted for other functions, like flight.

The reason the technoscientific method was coined as a new phrase is because the scientific and engineering methods are not independent of, but interact with, each other. The upper "fact finding" arrow exemplifies "engineering-based science." Because technology provides the means by which we interact with the world, technology is at the heart of fact-finding. The lower "form finding" arrow exemplifies "science-based engineering." Scientific knowledge enables us both to design forms likely to fulfill human-desired function and, especially, to eliminate forms likely *not* to fulfill human-desired function. And, to connect back to the previous section, as the scientific and engineering methods interact—with accumulated technological knowledge serving to advance scientific knowledge, and accumulated scientific knowledge serving to advance technological knowledge—knowledge translation between representations becomes important. Translating technology's embodied knowledge into connectionist and then symbolic knowledge enables the functional performance of technological forms to catalyze the creation of new scientific facts. And translating science's symbolic knowledge into connectionist and then embodied knowledge enables the explanations of science to catalyze the creation of new technological forms.

Importantly, we view this technoscientific method as being complete. Each of the six mechanisms is essential and there are no additional mechanisms. If one wanted to create artificial intelligences to advance technoscientific knowledge, they would need to execute *all* of these mechanisms, and they would also *only* need to execute these mechanisms.

Generate and Test Cycles

We turn now to how the various mechanisms of the technoscientific method are executed—through generate and test cycles whose aim is to

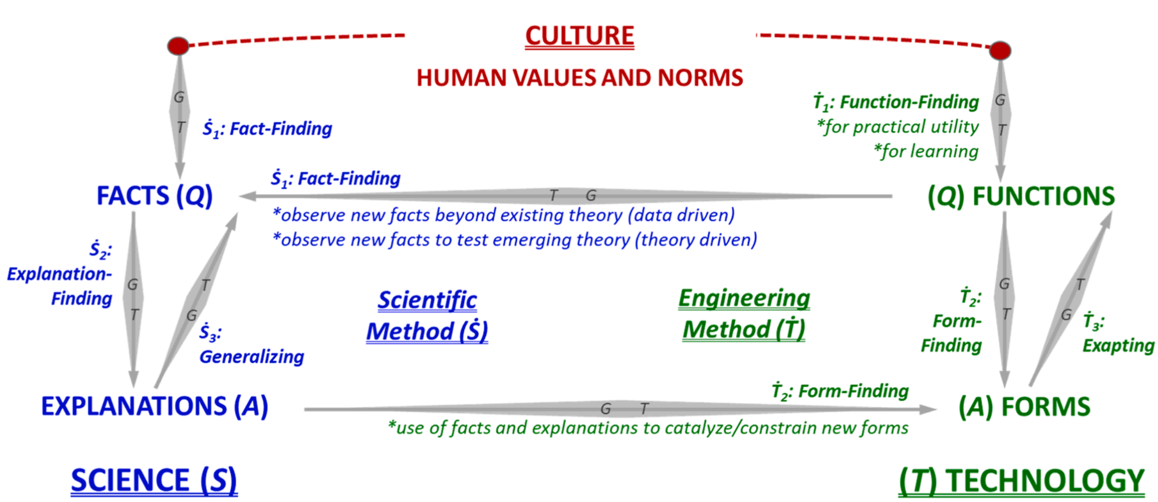

Fig. 4. The technoscientific method, comprised of interacting scientific (in blue on the left) and engineering (in green on the right) methods. The scientific method has three mechanisms: fact-finding, explanation-finding, and generalizing. The engineering method also has three mechanisms: function-finding, form-finding, and exapting. Each mechanism is represented by an arrow that widens as possibilities for knowledge advance are generated, then narrows as those possibilities are tested and filtered out. At the top in red is a placeholder for human culture—its values and norms—which interacts with the technoscientific method to co-determine what facts are of human-interest to measure and explain and what functions are of human interest to fulfill.

maximize the useful learning discussed earlier. These cycles are labeled *G* and *T* on each of the arrows in Fig. 4. To convey the sense in which the generate half of the cycles are of a divergent “generate new possibilities” nature, the arrow widens in that half; while to convey the sense in which the test half of the cycles are of a convergent “filter out the less desirable possibilities” nature, the arrow narrows in that half (Avina et al., 2018).

What an agent seeks when it executes these *G-T* cycles is outlined in Table 1. Each row of that table corresponds to a mechanism of the technoscientific method. Each column of that table corresponds to whether the agents are generating or testing possibilities for the finding of facts, explanations, functions, or forms.

The generation of new possibilities is represented by the second column of Table 1. The agent does not generate these new possibilities in a vacuum, but builds on what it knows—whether it is generating possible new facts by measurement, imagining possible new ways to explain facts, imagining possible new facts to measure, or imagining possible new human-desired functions or forms that might fulfill those functions. Importantly, because the world of possible measurements and ideas is so vast, the search for potentially useful measurements and ideas must be narrowed in some way, and one powerful way is heuristic and/or analogic reasoning (Hofstadter & Sander, 2013).

The testing of these new possibilities is represented by the third-through-fifth columns of Table 1, where we distinguish between various kinds of test.

The first distinction is between tests that are “virtual” (columns three and four) vs tests that are “real” (column five). Virtual tests are computational—either in human or artificial minds. These virtual tests are of course more efficient than real tests. As articulated by Karl Popper (Popper, 1968, p. 340): “Scientists try to eliminate their false theories, they try to let them die in their stead.” Biological evolution does not have this luxury—it tests the hard way, through possible death of actual living organisms. Still, virtual tests must be followed by real-world tests. The cumulative knowledge of society, whether human or artificial, will presumably always be limited. Ideas, even after they have passed their virtual test, must be tested for compatibility with, and utility in, the real world. Whether a new measurement technique does indeed yield desired

Table 1
Outline of criteria by which possibilities might be generated and tested (both virtually and in the real world) within each mechanism of the technoscientific method so as to maximize useful learning (implausible and/or plausible utility).

		GENERATE AND TEST CRITERIA			
		Generate Possibilities	Test (Virtually) the Possibilities for Implausible Utility (Research Flavor)	Test (Virtually) the Possibilities for Plausible Utility (Development Flavor)	Test (in the Real World) the Possibilities for Actual Utility
TECHNOSCIENTIFIC METHOD MECHANISMS	**$\dot{S}_1$: Fact-Finding (data driven)**	Measure possible facts: – that might go beyond existing theory	Test possible facts: – for maximization of the difference between the agent’s and society’s guesses as to whether they go beyond existing theory	Test possible facts: – for maximization of both the agent’s and society’s guesses as to whether they go beyond existing theory	Test the possible facts and explanations: – for the actual human interest of the facts that have been found, and – for how well the explanations explain the facts and "fan out" to explain multiple sets of facts
	$\dot{S}_2$: Explanation-Finding	Imagine possible explanations: – that might explain the current facts	Test possible explanations: – for maximization of the difference between the agent’s and society’s guesses as to whether they explain the current facts	Test possible explanations: – for maximization of both the agent’s and society’s guesses as to whether they explain the current facts	
	$\dot{S}_3$: Generalizing	Imagine possible additional facts: – that might be predicted by explanation of the current facts	Test possible additional facts: – for maximization of the difference between the agent’s and society’s reasonings as to whether they are predicted by explanations of the current facts	Test possible additional facts: – for maximization of both the agent’s and society’s reasonings as to whether they are predicted by the explanations of the current facts	
	$\dot{S}_1$: Fact-Finding (theory driven)	Measure facts: – that might be predicted by explanation of the current facts	Test measured facts: – for maximization of the difference between the agent’s and society’s guesses as to whether they are predicted by explanations of the current facts	Test measured facts: – for maximization of both the agent’s and society’s guesses as to whether they are predicted by explanations of the current facts	
	$\dot{T}_1$: Function-Finding (utilitarian)	Imagine possible functions: – that agents might desire	Test possible functions: – for maximization of the difference between the agent’s and society’s guesses as to whether agents will desire the function	Test possible functions: – for maximization of both the agent’s and society’s guesses as to whether agents will desire the function	Test possible functions and forms: – for the actual agent desirability of the functions that have been found, and – for how well the forms fulfill the function and "fan out" to fulfill multiple functions
	$\dot{T}_1$: Function-Finding (epistemic)	Imagine possible measurements: – that might generate facts that go beyond, or test, existing theory	Test possible measurements: – for maximization of the difference between the agent’s and society’s guesses as to whether they would generate facts that go beyond, or test, existing theory	Test possible measurements: – for maximization of both the agent’s and society’s guesses as to whether they would generate facts that go beyond, or test, existing theory	
	$\dot{T}_2$: Form-Finding	Imagine forms: – that might fulfill a desired function	Test possible forms: – for maximization of the difference between the agent’s and society’s guesses as to whether the form will fulfill the desired function	Test possible forms: – for maximization of both the agent’s and society’s guesses as to whether the form will fulfill the desired function	
	$\dot{T}_3$: Exapting	Imagine additional functions: – that might also be fulfillable by the current form	Test possible additional functions: – for maximization of the difference between the agent’s and society’s guesses as to whether the current form would fulfill the additional functions	Test possible additional functions: – for maximization of both the agent’s and society’s guesses as to whether the current form will fulfill the additional functions	

facts, whether a new form does indeed fulfill desired function—if just ideas they may be plausible, but only the real world can decide.

The second distinction is, within virtual tests, between tests for the implausible vs plausible utilities discussed earlier. The first kind of test, for implausible utility in column three, seeks possibilities that maximize the *difference* between the agent's and society's guesses as to the utility of the possibility. This is scenario 2 in Fig. 2(b), the scenario that maximizes the likelihood of surprise to society's cumulative knowledge, and thus to advance in that knowledge. The second kind of test, for plausible utility in column four, seeks possibilities that maximize *both* the agent's and society's guesses as to the utility of the possibility. This is scenario 1 in Fig. 2(b), the scenario that maximizes the likelihood that utility will be found. Both kinds of knowledge finding are important, so both kinds of tests are important.

Toward Artificial Scientists and Engineers

To give our emerging architecture more concreteness, in this final section we discuss examples of technoscientific knowledge advance as practiced by human intelligences and emerging artificial intelligences—by agent-based and autonomous (Green et al., 2022) scientists (King et al., 2009) and engineers (Sammut et al., 2015). We organize our examples around a full sequence of the technoscientific method. This method has no beginning or ending point, so in principle we could begin anywhere. We choose to begin with the engineering method, then to cross over to the scientific method. This is a classic ordering found throughout the history of technoscience, in which a newly invented form serves some human-desired function but defies existing scientific theory, triggering a hunt for a new theory: for example, the Nobel-Prize-winning invention of the blue light-emitting diode (Tsao et al., 2015).

Function-Finding (Utilitarian)

We start with the engineering method and, within the engineering method, with function finding of a utilitarian (practical) nature. Function finding is difficult, not just for an agent-based engineer, but also for a human engineer. As famously articulated by Steve Jobs (Isaacson, 2011, p. 567):

Some people say, "Give the customers what they want." But that's not my approach. Our job is to figure out what they're going to want before they do.

In other words, function-finding often requires imagination and going against common wisdom. Developing such intuition is a strong argument for artificial intelligence that is embodied. Embodiment is not just necessary for understanding the physical world in a direct way. It is also necessary for understanding the *human* world in a direct way. Humans learn what other humans desire in a host of ways inaccessible to artificial intelligence limited to second-hand descriptions. Just as "do as I say not as I do" doesn't work in teaching human children, it may not work in teaching artificial children.

Form-Finding

Once a function is found and assigned, the agent-based engineer's job is now to find a form that fulfills that function. Here, for brevity, we consider only physical forms (e.g., a hammer for driving nails), but, in the more general case, the form might also be algorithmic (e.g., methods for matrix multiplication, or for search, or for sorting).

The simplest method for form-finding is trial and error. Here, random variations or combinations of existing forms are tried and then tested in the real world for appropriateness. When the space of forms, or the mapping of forms to function, or both, are complex, however, this method of search is very inefficient. Thus, to aid their search, human and emerging agent-based engineers make use of engineering models of how those forms map to functions. These are models that exist in the conceptual, not the physical, world, and so can be implemented much more easily.

This modern form-finding loop is similar to trial and error, but with the slow outer loop of real test preceded by a fast inner loop of virtual test. For this fast inner loop (Bukkapatnam, 2023), the first step is an "inverse design" step that starts with a desired functionality and guesses a form that might fulfill that functionality (Zunger, 2018), and the second step is a "forward modeling" step that starts with a form and calculates whether its functionality is acceptable. If it is not, then a new guess is made, in a direction, in the conceptual space of forms, that might be more likely to yield a form with acceptable functionality.

This loop is powerful and when it is successful a new form has been discovered ("materials discovery" (Aspuru-Guzik & Persson, 2018; Pollice et al., 2021; Stach et al., 2021) or "chemicals discovery" (Coley et al., 2020a, 2020b)). In the nomenclature of the technoscientific method, however, it is not (or at least not necessarily) scientific discovery. Scientific knowledge might be used to great advantage in the engineering model that leads to materials or chemicals discovery (Rosen et al., 2021), but scientific knowledge itself is not necessarily being advanced. As will be discussed in later, however, scientific knowledge *could* be advanced—particularly when the form that is found is inconsistent with, and forces a rethinking of, existing scientific knowledge. The limits of scientific knowledge underscore the value of real-world test as the ultimate and most important test: we do not know what we do not know.

Nonetheless, for the sake of efficiently searching a complex space of possible forms, virtual testing is critical. And virtual testing in turn depends on the engineering model, whose form depends on how complicated the form-to-function relationship is.

At one extreme of complexity, if the form-to-function relationship is simple, the engineering model might be based on symbolic scientific knowledge—for example, the differential equations of Newton's laws specifying the ballistic behavior of an automated pizza thrower. Such an engineering model might still be non-trivial, with enormous computational resources necessary to implement, hence with room for artificial intelligence to accelerate. Modern examples include the use of computationally intensive density-functional theory (DFT) to calculate the functional properties of molecular forms (Rosen et al., 2021). Instead of DFT being used to calculate the functional properties of every guessed form, it is used to calculate properties for a representative and smaller set of forms, which then are used to train a machine learning model, which then becomes the less-computationally-intensive engineering model. In the language of knowledge representations, this is "instinctivizing the symbolic" (De Regt, 2017).

At the other extreme of complexity, if the form-to-function relationship is complex, the engineering model must be based on data—-forms that are embodied in the real world and whose functionality is tested in the real world. This is the domain of big data: data taken over the entire space of forms and their observed (measured) functionalities. Often this is more data—multi-modal and of high dimensionality—than any human intelligence could make sense of, so, again, there is room for artificial intelligence to accelerate. Machine learning models can be trained on this form-to-function data, then subsequently used as the engineering model that does form-to-function inference.

Between the two extremes of complexity, scientific computing and machine learning can be blended (SciML = Scientific Machine Learning) (DOE/SC/ASCR, 2019; Feddema, 2019). Scientific or other human knowledge (Gil et al., 2019) can be used to enable a machine learning model to converge more quickly as it is being trained on data (PINN = physics-informed neural networks (Raissi et al., 2019)).

Exapting

Once a form has been found to fulfill a desired function, and once it has been embodied, the human or agent-based engineer's immediate, assigned, job is done. But, to maximize technoscientific advance, advantage can be taken of the opportunity for exapting. The human or artificial engineer has direct, unique insight into the form that it has created, hence into its potential for unintended functionality that could also be human or agent desired even if not specified at the outset of form-finding. The microwave oven is a classic example (Osepchuk, 1984): Percy Spencer leaning against an open microwave waveguide, noticing that a candy bar melted in his pocket, then putting two and two together to imagine microwaves as a source of energy for cooking and not just for radar. This example underscores the importance of being embodied in, and interacting directly with, the physical world both for serendipitous inspiration, as well as for real-world validation of the potential, of new ideas. Though current large-language models are proving to be excellent at the divergent thinking that puts such two-and-two's together (Guzik et al., 2023), they are not yet excellent at the convergent thinking necessary to validate the real-world correctness, or in this example potential, of ideas.

Importantly, exapting does not need to wait until a form has been found, embodied, and entered into real-world use. It can also occur during the process of form-finding. It is a common research occurrence to explore forms for some intended functionality, but then to notice that, even if those forms appear not to fulfill the intended functionality, they might fulfill some other functionality. 3M's Post-It Notes are a classic example (Nayak & Ketteringham, 1986): Spencer Silver finding an adhesive that was much weaker than what he was looking for but with the unusual property of removability, hence finding utility as notes that could be attached then easily removed from books and other surfaces. This example underscores the importance of building blocks or intermediate forms that do not make it out into embodied societal knowledge but, if kept in mind, can be useful for future functionalities. One might imagine agent-based engineers, much like human engineers, collecting libraries of such building blocks for future use (Wang, Xie, et al., 2023), in a manner similar to hindsight experience replay (Andrychowicz, 2017).

Fact-Finding (Data-Driven)

We turn now to the mechanisms of the scientific method, and begin with fact-finding of an open-ended data-driven nature—the finding of facts that go beyond existing theory. Here we distinguish between two possibilities.

The first possibility is serendipitous and follows from form-finding: when a form is found whose functional properties differ from what conventional scientific wisdom would predict. Heavier-than-air flight, for example, becomes a new scientific fact that goes against conventional scientific wisdom hence spawns a hunt for new scientific explanations. In a sense, these new forms and their functional properties were exapted to serve as facts begging for scientific explanation (Bukkapatnam, 2023).

The second possibility is planned and also follows from form-finding. When new forms with new functionalities are discovered—new telescopes, new microscopes, new oscilloscopes, new materials synthesis techniques—these open new windows into the world. It is natural to look through these new windows even without knowing what one will find, indeed *because* one does not know what one will find. When Galileo turned his new telescope outwards to the sky and when Leeuwenhoek turned his new microscope inwards to small objects, they did indeed find new facts—the existence of the moons of Jupiter and the existence of bacteria and other microscopic life. In modern times, particle accelerators at ever higher energies are a new window into more and more strongly bound particles and sub-particles.

Explanation-Finding

With new facts found, explanation-finding comes next. And, by explanation-finding we do not mean ultimate explanations, we mean the entire range of explanations from the shallowest to the deepest. At the shallow end, we may merely be finding simple patterns in our observations of nature; at the deep end, we may be finding deeper patterns in those shallower patterns (Root-Bernstein, 1991). As perhaps first articulated by Newton and perhaps the philosophy that most untethered science from religion, it is sufficient for one's explanation to explain, even if one's explanation does not yet have an explanation. The then-unexplainability of action at a distance did not prevent gravity from explaining the motion of planets.

That said, pattern finding is ubiquitous not just in science but also in engineering: the engineering models discussed earlier are essentially the finding of patterns in form-function relationships. What distinguishes this kind of pattern finding with an engineering purpose from pattern finding with a scientific purpose?

The distinction is not, as some might claim, in the depth of the pattern finding, though scientific pattern finding may sometimes be deeper. The distinction has to do with the functional purpose that drives the specificity vs generality of the pattern finding. Engineering pattern-finding, or engineering modeling, must predict specific function and performance. It cannot afford to simplify beyond the needs of the desired function so as to facilitate generalization to other functions. It must be accurate in relevant real-life situations because the engineered forms must perform in those real-life situations. Simplification in engineering modeling most often occurs because those details are irrelevant to the task at hand. If you can only affordably purchase pipes with diameters of 2" or 4", there is no need to model the flow in hypothetical pipes with other diameters, or diameters precise to several decimal places. Scientific modeling, in contrast, might very well simplify so as to enable a particular phenomenon, like gravity as a force on earth, to generalize to other situations, like the orbits of planets around the sun.

Most importantly, scientific explanation finding does not just follow from the facts at hand, but builds on explanations that have previously explained other facts, sometimes from knowledge domains far from the current knowledge domain. Imaginative leaps are thus required, leaps which benefit from analogical mappings between domains that are conceptually and structurally similar. Large language models that can "see" similarities between two different fields (O'Brien et al., 2023) provide a recent example of such analogical inference. From this perspective, an artificial agent could exceed a human counterpart not so much because it might operate faster and more efficiently, though it might also do that (MacLeod et al., 2023), but because the imaginative analogical leaps it makes could span a much wider range of knowledge domains.

Generalizing

As just discussed, generalization is key to scientific explanations. It is generalization, using explanations of one set of facts to also explain additional facts (Feynman, 1974), that creates the seamless web of scientific facts and explanations—the dense interconnections that gives strength to the entire web (Anderson, 2001). Also, the more "different" the new predicted facts are from the original facts, the more powerful the generalization. That $E=mc^2$, originally intended to explain the constancy of the speed of light, also explains the release of energy upon mass change, is an extremely strong argument for its generality as a law of nature.

Moreover, just as in the engineering method the interactivity inherent in the form-finding/exapting loop is powerful, so in the scientific method the interactivity inherent in the explanation-finding/generalizing loop is powerful. An explanation, once found, is immediately either generalized to other facts that are consistent, or updated in the face of other facts that are inconsistent. Humans at all stages of life, even children (Gopnik, 2012), employ this theory/experiment strategy to make sense of new facts and situations they are confronted by.

The form this strategy takes, though, may differ depending on stage of life. In an early stage, when theory is underdeveloped, agents may rely more heavily on experiment and pure induction from experiment to develop theory. At an intermediate stage, when theory is more developed, agents may rely more equally on theory and experiment: using theory to predict what to expect from experiment and paying selective attention to mis-predictions to update theory (Deutsch, 2011). At a later stage, when theory is perhaps overdeveloped, agents may rely too heavily on theory, misperceiving or misinterpreting experiment to force-fit to theory rather than updating theory. The sweet spot is the intermediate stage, with the greatest potential for few-shot learning (Wang et al., 2020) rather than many-shot learning or absence of learning, and with the greatest potential for continuous lifelong learning (Chaudhry et al., 2018; Parisi et al., 2019).

Function- and Form-Finding (Epistemic), and Fact-Finding (Theory Driven)

It is one thing to have generalized an explanation to the prediction of possible new facts. It is another thing to devise an experiment or a measurement to observe those possible new facts. This brings us full circle back to the engineering method. We now have a new human-desired function, this time of an epistemic nature: observe the predicted new fact so as to advance scientific knowledge by making more plausible or implausible the emerging explanation that predicts that new fact.

Closing this engineering loop is tricky. There are so many different kinds of experimental apparatuses, so many different combinations of these. One needs an enormous library, even in very narrow specialized areas of research, to make progress. Often human experimenters suffer from a "search where the lamp shines" limitation of what tools are readily available. In principle, artificial experimenters can access a much wider range of tools, but they must have the equivalent of the human eyes and hands to which most tools have been tailored. Ultimately, though they might build their own ecosystem of tools geared to them rather than to humans (Maffettone et al., 2023).

Recapitulation

Our architecture weaves together perspectives from a wide range of disciplines. Here, we recapitulate the most salient of these perspectives:

First, technoscientific knowledge advance and accumulation is a multi-agent coopetition, in which agents compete to advance, but cooperate so that society can accumulate, technoscientific knowledge.

Second, scientific facts and explanations, and technological functions and forms, are the fundamental building blocks of technoscientific knowledge. Their interconnections within the "seamless webs of knowledge" are measures of their utility and importance. Thus, exaptation in the engineering method and generalization in the scientific method are both key, as they create such interconnections.

Third, the technoscientific method represents the fundamental mechanisms by which science and technology advance; with the scientific and engineering sub-methods both similar but very different (e.g., scientific modeling $\neq$ engineering modeling).

Fourth, generation and (virtual + real) testing of new nuggets of knowledge for implausible and plausible utility are the fundamental means of executing the mechanisms of the technoscientific method, with implausible utility being the epitome of societal-level creative discovery.

Fifth, symbolic and embodied knowledge representations are key to societal accumulation, while a connectionist knowledge representation is key to agential advance, of technoscientific knowledge.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

D.C. Crowder has patent AI-Enhanced Codesign for Circuit and System Design pending to Pending, not yet issued. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. The work was in part supported by Sandia National Labs' Laboratory Directed Research and Development program (#230710) and in part by the Center for Integrated Nanotechnologies, an Office of Science user facility operated for the U.S. Department of Energy. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the US Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

References

Anderson, P. W. (1972). More is different. *Science, 177*(4047), 393–396.
Anderson, P. W. (2001). Science: A 'dappled world' or a 'seamless web'? *Stud. Hist. Phil. Mod. Phys., 32*(3), 487–494.
Andriani, P., & Carignani, G. (2014). Modular exaptation: a missing link in the synthesis of artificial form. *Research Policy, 43*(9), 1608–1620.
Arthur, W. B. (2009). *The nature of technology: What it is and how it evolves*. Simon and Schuster.
Aspuru-Guzik, A., & Persson, K. (2018). Materials Acceleration Platform: Accelerating Advanced Energy Materials Discovery by Integrating High-Throughput Methods and Artificial Intelligence. *Report of the Clean Energy Materials Innovation Challenge Workshop, sponsored by the Mexican Ministry of Energy (SENER), the U.S. Department of Energy (US DOE), and the Canadian Institute for Advanced Research (CIFAR)*.
Avina, G. E., Schunn, C. D., Odumoso, T., Silva, A. R., Bauer, T. L., Crabtree, G. W., Johnson, C. M., Picraux, S. T., Sawyer, R. K., Schneider, R. P., Sun, R., Feist, G. J., Narayanamurti, V., & Tsao, J. Y. (2018). The Art of Research: A Divergent/Convergent Thinking Framework and Opportunities for Science-Based Approaches. In J. Y. Tsao, Eswaran Subramanian, & Toluwalogo Odumoso (Eds.), *Engineering a Better Future*. Springer.
Bengio, Y., Lecun, Y., & Hinton, G. (2021). Deep learning for AI. *Communications of the ACM, 64*(7), 58–65.
Boden, M. A. (2005). What is creativity?. *Creativity in human evolution and prehistory* (pp. 27–55). Routledge.
Booch, G., Fabiano, F., Horesh, L., Kate, K., Lenchner, J., Linck, N., Loreggia, A., Murgesan, K., Mattei, N., & Rossi, F. (2021). Thinking fast and slow in AI. In *Paper presented at the Proceedings of the AAAI Conference on Artificial Intelligence*.
Bukkapatnam, S. T. (2023). Autonomous materials discovery and manufacturing (AMDM): A review and perspectives. *IISE Transactions, 55*(1), 75–93.
Chaudhry, A., Ranzato, M. A., Rohrbach, M., & Elhoseiny, M. (2018). Efficient lifelong learning with A-GEM. *arXiv preprint arXiv:1812.00420*.
Clark, A., & Chalmers, D. (1998). *The extended mind. analysis, 58*(1), 7–19.
Coley, C. W., Eyke, N. S., & Jensen, K. F. (2020a). Autonomous discovery in the chemical sciences part I: Progress. *Angewandte Chemie International Edition, 59*(51), 22858–22893.
Coley, C. W., Eyke, N. S., & Jensen, K. F. (2020b). Autonomous discovery in the chemical sciences part II: outlook. *Angewandte Chemie International Edition, 59*(52), 23414–23436.
Davis, E., & Aaronson, S. (2023). Testing GPT-4 with Wolfram Alpha and Code Interpreter plug-ins on math and science problems. *arXiv preprint arXiv:2308.05713*.
De Regt, H. W. (2017). *Understanding scientific understanding*. Oxford University Press.
Deutsch, D. (2011). *The beginning of infinity: Explanations that transform the world*. Penguin UK.
DOE/SC/ASCR. (2019). *Basic Research Needs Workshop for Scientific Machine Learning: Core Technologies for Artificial Intelligence*. U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research. *February 10, 2019*.
Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research Policy, 11*(3), 147–162.
Eagleman, D. (2011). *Incognito: The Secret Lives of the Brain*. New York City: Pantheon.
Feddema, J. (2019). An Overview of Machine Learning for Scientific and High Performance Computing at Sandia. *Sandia National Laboratories. SAND2019-5645PE*.
Feynman, R. (1974). *Cargo cult science: CalTech 1974*. Commencement Address.
Funk, R. J., & Owen-Smith, J. (2017). A dynamic network measure of technological change. *Management Science, 63*(3), 791–817.
Gibson, J. J. (1977). The theory of affordances. In James J Gibson (Ed.), *The Ecological Approach to Visual Perception*. Hilldale, USA: Psychology Press, Taylor & Francis.
Gil, Y., Honaker, J., Gupta, S., Ma, Y., D'Orazio, V., Garijo, D., Gadewar, S., Yang, Q., & Jahanshad, N. (2019). Towards human-guided machine learning. In *Paper presented at the Proceedings of the 24th International Conference on Intelligent User Interfaces*.
Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science, 337*(6102), 1623–1627.

Green, M. L., Maruyama, B., & Schrier, J. (2022). Autonomous (AI-driven) materials science. *Applied Physics Reviews, 9*(3).
Guzik, E. E., Byrge, C., & Gilde, C. (2023). The originality of machines: AI takes the Torrance Test. *Journal of Creativity, 33*, Article 100065.
Hadjimichael, D. (2023). Beyond Tacit Knowledge: How Michael Polanyi's Theory of Knowledge Illuminates Theory Development in Organizational Research. *Academy of management review. in press*.
Harré, R. (1981). *Great scientific experiments: Twenty experiments that changed our view of the world*. Phaedon Press Limited.
Hayek, F. A. (1945). The use of knowledge in society. *The American economic review, 35* (4), 519–530.
Hennessey, B., Moran, S., Altringer, B., & Amabile, T. M. (2015). Extrinsic and intrinsic motivation. In Cary L. Cooper (Ed.), *Wiley Encyclopedia of Management* (pp. 1–4). John Wiley & Sons.
Henrich, J. (2015). *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.
Hofstadter, D., & Sander, E. (2013). *Surfaces and essences: Analogy as the fuel and fire of thinking*. Basic Books.
Isaacson, W. (2011). *Steve Jobs: The Exclusive Biography*. Simon & Schuster.
James, C. D., Aimone, J. B., Miner, N. E., Vineyard, C. M., Rothganger, F. H., Carlson, K. D., Mulder, S. A., Draelos, T. J., Faust, A., & Marinella, M. J. (2017). A historical survey of algorithms and hardware architectures for neural-inspired and neuromorphic computing applications. *Biologically Inspired Cognitive Architectures, 19*, 49–64.
Johnson, M. (2015). Embodied understanding. *Frontiers in psychology, 6*, 875.
Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.
Kambhampati, S. (2021). Polanyi's revenge and AI's new romance with tacit knowledge. *Communications of the ACM, 64*(2), 31–32.
King, R. D., Rowland, J., Aubrey, W., Liakata, M., Markham, M., Soldatova, L. N., Whelan, K. E., Clare, A., Young, M., & Sparkes, A. (2009). The robot scientist Adam. *Computer, 42*(8), 46–54.
Krenn, M., Pollice, R., Guo, S. Y., Aldeghi, M., Cervera-Lierta, A., Friederich, P., Gomes, G.d. P., Häse, F., Jinich, A., & Nigam, A. (2022). On scientific understanding with artificial intelligence. *Nature Reviews Physics, 4*(12), 761–769.
Kuhn, T. S. (1962). *The Structure of Scientific Revolutions* (4th ed.). University of Chicago Press. *Fourth edition*(4th edition published in 2012).
Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Harvard University Press.
LeCun, Y. (2022). *A Path Towards Autonomous Machine Intelligence Version 0.9, 2*, 2022-06-27*(Unpublished report*.
Leibel, C., & Bornmann, L. (2023). What do we know about the disruption indicator in scientometrics? An overview of the literature. *arXiv preprint arXiv:2308.02383*.
MacLeod, B. P., Parlane, F. G., & Berlinguette, C. P. (2023). How to build an effective self-driving laboratory. *Mrs Bulletin, 48*(2), 173–178.
Maffettone, P. M., Friederich, P., Baird, S. G., Blaiszik, B., Brown, K. A., Campbell, S. I., Cohen, O. A., Collins, T., Davis, R. L., & Foster, I. T. (2023). What is missing in autonomous discovery: Open challenges for the community. *arXiv preprint arXiv: 2304.11120*.
Marcus, G. F. (2003). *The algebraic mind: Integrating connectionism and cognitive science*. MIT press.
Merchant, B. (2017). *The one device: The secret history of the iPhone*. Random House.
Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago Press.
Narayanamurti, V., & Tsao, J. Y. (2021). *The Genesis of Technoscientific Revolutions: Rethinking the Nature and Nurture of Research*. Harvard University Press.
Nayak, P. R., & Ketteringham, J. M. (1986). *Breakthroughs!* New York: Rawson Associates.
Newell, A. (1994). *Unified theories of cognition*. Harvard University Press.
Newton, I. (1675). *Letter from Sir Isaac Newton to Robert Hooke*. Historical Society of Pennsylvania.
O'Brien, T., Stremmel, J., Pio-Lopez, L., McMillen, P., Rasmussen-Ivey, C., & Levin, M. (2023). Machine Learning for Hypothesis Generation in Biology and Medicine: Exploring the latent space of neuroscience and developmental bioelectricity. *Preprint*.
Osepchuk, J. M. (1984). A history of microwave heating applications. *IEEE Transactions on Microwave Theory and Techniques, 32*(9), 1200–1224.
Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual lifelong learning with neural networks: A review. *Neural Networks, 113*, 54–71.
Polanyi, M. (1966). *The tacit dimension: The*. University of Chicago Press.
Pollice, R., dos Passos Gomes, G., Aldeghi, M., Hickman, R. J., Krenn, M., Lavigne, C., Lindner-D'Addario, M., Nigam, A., Ser, C. T., & Yao, Z. (2021). Data-driven strategies for accelerated materials design. *Accounts of chemical research, 54*(4), 849–860.
Popper, K. R. (1968). *Epistemology without a knowing subject Studies in Logic and the Foundations of Mathematics, 52* pp. 333–373). Elsevier.
Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics, 378*, 686–707.
Root-Bernstein, R. (1991). *Discovering*. Harvard University Press.
Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explainable machine learning for scientific insights and discoveries. *IEEE Access, 8*, 42200–42216.
Rosen, A. S., Iyer, S. M., Ray, D., Yao, Z., Aspuru-Guzik, A., Gagliardi, L., Notestein, J. M., & Snurr, R. Q. (2021). Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. *Matter, 4*(5), 1578–1597.
Sammut, C., Sheh, R., Haber, A., & Wicaksono, H. (2015). August 20-22. In *The Robot Engineer. Paper presented at the 2015 International Conference on Inductive Logic Programming*. Kyoto, Japan: Kyoto University.
Schleder, G. R., Padilha, A. C., Acosta, C. M., Costa, M., & Fazzio, A. (2019). From DFT to machine learning: recent approaches to materials science–a review. *Journal of Physics: Materials, 2*(3), Article 032001.
Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American philosophical society, 106*(6), 467–482.
Simonton, D. K. (2013). What is a creative idea? Little-c versus Big-C creativity. *Handbook of research on creativity* (pp. 69–83).
Simonton, D. K. (2018). Defining Creativity: Don't We Also Need to Define What Is Not Creative? *The Journal of Creative Behavior, 52*(1), 80–90.
Stach, E., DeCost, B., Kusne, A. G., Hattrick-Simpers, J., Brown, K. A., Reyes, K. G., Schrier, J., Billinge, S., Buonassisi, T., & Foster, I. (2021). Autonomous experimentation systems for materials development: A community perspective. *Matter, 4*(9), 2702–2726.
Stevens, R., Taylor, V., Nichols, J., Maccabe, A. B., Yelick, K., & Brown, D. (2019). *AI for Science: Report on the Department of Energy (DOE) Town Halls on Artificial Intelligence (AI) for Science*. U.S. Department of Energy, Office of Science.
Tanaka, S. (2011). The notion of embodied knowledge. *Theoretical psychology: Global transformations and challenges* (pp. 149–157).
Economist, The (2023). Could AI transform science itself. *The Economist*.
Thiel, P., & Masters, B. (2014). *Zero to one: Notes on startups, or how to build the future: Currency*.
Tsao, J. Y., Han, J., Haitz, R. H., & Pattison, P. M. (2015). The Blue LED Nobel Prize: Historical context, current scientific understanding, human benefit. *Annalen der Physik, 527*(5-6).
Tsao, J. Y., Ting, C. L., & Johnson, C. M. (2019). Creative Outcome as Implausible Utility. *Review of General Psychology, 23*, 279.
Valiant, L. G. (2013). *Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World*. Basic Books.
Viereck, G. S. (1929). What life means to Einstein: an interview. *The Saturday Evening Post, 26*, 17.
Wang, G., Xie, Y., Jiang, Y., Mandlekar, A., Xiao, C., Zhu, Y., Fan, L., & Anandkumar, A. (2023). Voyager: An open-ended embodied agent with large language models. *arXiv preprint arXiv:2305.16291*.
Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., Chandak, P., Liu, S., Van Katwyk, P., & Deac, A. (2023). Scientific discovery in the age of artificial intelligence. *Nature, 620*(7972), 47–60.
Wang, Y., Yao, Q., Kwok, J. T., & Ni, L. M. (2020). Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys (CSUR), 53*(3), 1–34.
Watson, R., & Levin, M. (2023). The collective intelligence of evolution and development. *Collective Intelligence, 2*(2), Article 26339137231168355.
Wolpert, D. H., & Tumer, K. (1999). An introduction to collective intelligence. *arXiv preprint cs/9908014*.
Zunger, A. (2018). Inverse design in search of materials with target functionalities. *Nature Reviews Chemistry, 2*(4), 0121.