

What is the Value of One(Few)-Shooting Moonshots?*

Akshit Kumar[†]

Preliminary Draft. Comments Welcome.

Abstract

Good research is not only about producing breakthroughs but also communicating the methods and connections that enable further advances. AI can accelerate discovery while revealing little of the broader knowledge generated during its search. Human research may proceed more slowly, yet leave a richer foundation for future work. Which of these serves society better in the long run? We study this question in a model of parallel exploration on a two-dimensional grid of ideas. Humans communicate their explored map, while AI communicates one successful path to its deepest result. Social value is $\lambda D^\beta + \mu|P|$, where D is the deepest discovery, P is the set of communicated concepts, $\beta \geq 1$ measures the value of depth, and $\lambda, \mu > 0$ weight breakthroughs and public knowledge. AI has a multiplicative search advantage $A \geq 1$ and a difficulty-exponent advantage $a \geq 1$. In the quadratic benchmark, $\beta = 2$ and $a = 1$, the ratio of expected AI to human value converges to $[\lambda/(\lambda + \mu c_\alpha)]A^{2/(\alpha+1)}$, where $\alpha > 0$ governs increasing research difficulty and c_α measures the area of the human knowledge map relative to squared depth. This yields an explicit capability threshold. More generally, a multiplicative advantage changes the scale of discovery, while an advantage in the difficulty exponent changes its growth rate. We characterize when human breadth prevails, when deeper AI discoveries eventually dominate, and when social valuations decide the comparison. The results help clarify why communities can reasonably reach different judgments about the value of one-shooting difficult research problems.

1 Introduction

Human beings have been dethroned from the top of the theorem-proving food chain and no longer occupy an uncontested position at the top of the problem-solving hierarchy. OpenAI’s October 2026 release of more than 700 mathematical manuscripts demonstrated that frontier AI models are capable of producing interesting and nontrivial mathematics, an ability once largely confined to a select group of *human* mathematicians. The release, which some called a *proof dump*, reported advances on long-standing open problems in mathematics and theoretical computer science (OpenAI, 2026). Some luminaries and thought leaders in the mathematical community have expressed concern about the way OpenAI released these results. Alongside correctness, these concerns involve the difficulty of assimilating an argument that seems to have been parachuted into their field without the related explanations, failed approaches, and connections that make it part of a field’s working knowledge. A September 2026 declaration by a group of mathematicians emphasizes that major open problems motivate the development of concepts and techniques, whose value emerges through

*The conceptualization and model in this paper is entirely that of the human author. GPT-6 Astra proved the main results, found relevant connection to literature and paraphrased parts of the paper.

[†]University of Toronto, Email: akshit.kumar@utoronto.ca

exposition, discussion, and subsequent work (Math and AI, 2026). The concern is that such large scale unfettered production of solutions can become detached from the development and transmission of understanding.

A difficult problem in mathematics is often more about the journey than the destination. In trying to solve it, researchers test ideas, discover obstructions, develop methods, and learn which neighboring questions may be worth pursuing. Some approaches fail to solve the original problem and still reveal connections that later become useful elsewhere. This knowledge is communicated through papers, conferences, seminars, collaboration, and the training of subsequent generations. From this perspective, maybe the real treasure is the connections made along the way. An AI system that makes the connections internally and releases a finished, often extremely hard to parse, proof may convey the conclusion while leaving much of the intellectual map created in reaching it inaccessible.

This concern in fact predates modern AI. In his now celebrated essay, *On proof and progress in mathematics*, Thurston (1994) asks, “How do mathematicians advance human understanding of mathematics?” His account emphasizes the communication of methods and ways of thinking, and the collective work that makes results usable by others. AI sharpens an existing distinction between establishing a conclusion and developing the knowledge around it. Tao (2026a) describes the strain between rapid proof production and slower assimilation as *proof indigestion*. He writes:

“A proof that no human can properly explain should be viewed as incomplete, even if it has been formally verified.”

Imagine that a super sophisticated AI model is able to resolve the P versus NP problem. Establishing the answer would settle a fundamental question, while understanding the techniques behind it could enable other advances in algorithms or complexity theory. These gains depend on what the proof supplies: even a proof of $P = NP$ may fail to yield practically useful algorithms. The conclusion and the methods it makes available are therefore distinct sources of value, and different mathematical communities may attach different weights to them.

The same tension arises when individuals use AI to conduct research. Democratizing intelligence enables non-experts to tackle difficult questions through one or a few prompts and release proposed solutions as preprints. Even when these users verify a solution and understand how each step follows from the preceding ones, their accounts may convey little about how the key ideas arose, which natural approaches failed, or what neighboring questions deserve attention. Communicating these details can help others develop new methods and pursue further discoveries. A human-authored account that follows only the AI’s successful argument may therefore leave much of the surrounding intellectual map unexplored or uncommunicated.

While mathematics is one of the first disciplines which is grappling with this existential crisis of sorts, it is certainly not going to be the last. These questions extend naturally to physics, chemistry, biology, and medicine to name a few, where AI may also help resolve long-standing problems. Scientists in these fields have strong reasons to understand the processes behind a discovery, including the possibility of improving it or applying its methods elsewhere. At the same time, the direct value of a deep discovery can be revolutionary. Consider an AI-discovered treatment that is *independently and rigorously* established to be safe and effective against a particular cancer or another serious disease — an analogue of resolving the P versus NP problem. The benefit to patients could be substantial even while the scientific community has only a partial understanding of the research process that produced it. Further understanding could support better treatments and advances against other diseases. Yet a sufficiently valuable, usable outcome could compensate for a narrower account of the surrounding discoveries.

This dichotomy then begs the question, what is really the value of one(few)-shotting these moonshots? Should an organization like the National Science Foundation favor research that solves extremely deep and challenging problems while conveying little of how it got there? Or should it favor a slower process whose participants leave a broader collection of ideas and techniques for future researchers?

The reality, of course, will lie somewhere between these extremes. People are already using AI to deepen their understanding and communicating at least some of what they learn, though perhaps less than we might hope. Our hope is that, by developing some understanding of what happens at the extremes, we can begin to see how to chart a course through the middle.

This paper attempts to provide some formalism to this debate about what the future of doing and communicating research might look like, and which approaches might benefit society in the long run. We are interested in understanding what each approach contributes, what we might lose along the way, and how that balance changes as AI becomes more capable. We recognize that a simple model can capture only part of this story. Still, we hope it offers a useful starting point for thinking through choices that many researchers are already confronting: how much to rely on AI, how deeply to engage with the ideas it produces, and how to communicate what they learn so that others can build on it.

Model. We imagine research as gradually finding one’s way through a landscape of ideas. We represent this landscape by a two-dimensional grid, with our starting knowledge at the origin. Each node represents a concept, and establishing a connection to a neighboring node allows us to build on what we already know. A newly discovered concept opens several possible directions for further work. Some attempts succeed quickly, while others require repeated effort. As research moves farther from its starting point, connections become harder to establish, with a parameter $\alpha > 0$ governing how quickly this difficulty increases.

Both human and AI researchers explore in this way. They can pursue many connections in parallel, revisit unsuccessful attempts, and build on everything they have discovered. We allow AI to be better in two ways. A multiplier $A \geq 1$ raises its probability of establishing a connection. A second parameter $a \geq 1$ reduces the exponent of research difficulty from α to α/a . This distinction matters to the story we want to tell. AI might be faster at the same kinds of investigations that humans pursue. It might also become increasingly better at handling the difficulties that arise as research moves into unfamiliar territory.

In our human research mode, researchers make their explored map available through papers, conference and workshop presentations, seminars, and informal conversations. In the AI research mode, the community receives a single successful path to the deepest result. The AI may have explored many routes and made repeated attempts internally. Here, *one-shotting* refers to the experience of receiving the completed solution, with its successful route communicated. These are deliberately extreme accounts of communication. They help us isolate a concern that can also arise when people use AI and share only the argument that worked.

We value research through two contributions. Reaching depth D delivers a breakthrough worth λD^β , where $\beta \geq 1$ describes how strongly society rewards deeper achievements. Communicating a set P of concepts creates public knowledge worth $\mu|P|$. A method developed on an unsuccessful route can still help someone approach another question. Counting communicated concepts is a simple way to recognize these additional opportunities. Each distinct concept receives the same credit, whether it came from a human or an AI, and repeated explanations of the same concept do

not increase its count. Our social value is therefore

$$U = \lambda D^\beta + \mu |P|.$$

The weights λ and μ allow communities to attach different importance to a usable breakthrough and the broader knowledge available for future work. We recognize how much this simple count leaves out, including differences in the usefulness of concepts and the quality of their explanation. Its appeal is that it makes the comparison transparent.

Main findings. The main finding of our model is a long-run comparison between the social value of AI-led and human-led research. Let $R(t, A, a) = \mathbb{E}[U_A(t, A, a)]/\mathbb{E}[U_H(t)]$ denote their ratio of expected values at time t . The clearest benchmark takes $\beta = 2$ and $a = 1$. Human knowledge then expands over an area whose value has the same order as the squared depth of the best breakthrough. For every fixed $A \geq 1$, we prove that

$$\lim_{t \rightarrow \infty} R(t, A, 1) = \frac{\lambda}{\lambda + \mu c_\alpha} A^{2/(\alpha+1)}, \quad A^* = \left(1 + \frac{\mu}{\lambda} c_\alpha\right)^{(\alpha+1)/2}. \quad (1)$$

The coefficient c_α measures the limiting area of the human community’s public map relative to squared maximum depth. Its value is bounded explicitly even though the exact exploration shape is unknown.

What we find useful about this expression is how clearly it separates the forces involved. The factor $A^{2/(\alpha+1)}$ measures AI’s advantage in breakthrough value. The fraction $\lambda/(\lambda + \mu c_\alpha)$ captures the additional contribution of human public knowledge. The concepts along AI’s communicated path also have value. In this benchmark, that path grows more slowly than squared depth, so its contribution becomes a vanishing fraction of AI’s total value. Human research eventually generates greater expected value when $A < A^*$, and AI research does so when $A > A^*$.

The general model lets us ask how robust this trade-off is. With a purely multiplicative advantage, human breadth eventually prevails for $1 \leq \beta < 2$, while every $A > 1$ gives AI a limiting advantage when $\beta > 2$. Changing the difficulty exponent has a stronger effect. Any $a > 1$ makes AI’s value grow faster when $\beta \geq 2$. For $1 \leq \beta < 2$, there is a finite critical exponent advantage precisely when $\beta(\alpha + 1) > 2$. In that case,

$$a_c = \frac{2\alpha}{\beta(\alpha + 1) - 2}$$

separates eventual human dominance from eventual AI dominance. At the boundary, the multiplier and social valuations decide the comparison. When $\beta(\alpha + 1) \leq 2$, human breadth prevails for every fixed finite pair (A, a) .

How we interpret these findings depends on what we care about. The weights λ and μ may differ across disciplines and communities, including among mathematicians. For mathematicians who place particular value on developing connections and opening directions for others to explore, one-shotting difficult problems, and the OpenAI stunt in particular, may seem pernicious. A completed solution can leave too much of the knowledge generated along the way inaccessible. In the quadratic benchmark, a larger weight on public knowledge raises the capability advantage needed to compensate for that loss.

We may feel differently when a discovery offers the prospect of saving lives or substantially improving their quality. If AI produces an effective treatment for a serious disease, society may attach an enormous value to the achievement. A larger λ relative to μ captures that possibility. Greater convexity in the value of depth, represented by β , also makes distant discoveries more

consequential. And if AI changes how rapidly difficulty grows with depth, its contribution can eventually outgrow the broader knowledge produced by human research.

These findings give us a more qualified picture of what increasing AI capability might mean. The comparison depends on the value of deeper results, the breadth of communication, and the kind of advantage AI has. We hope these distinctions help make disagreements about the future of research a little more explicit.

Related literature. The closest economic antecedent is *A Quest for Knowledge* by [Carnehl and Schneider \(2025\)](#). They study how existing knowledge shapes both the usefulness of new discoveries and researchers’ choices about which questions to pursue. Their analysis distinguishes expanding the knowledge frontier from filling gaps between established discoveries. A particularly relevant result is that a distant moonshot can benefit society by encouraging subsequent researchers to connect it to earlier knowledge. This gives a formal account of something central to our motivation: the value of a breakthrough depends partly on the research that develops around it. We build on this concern by examining how much of the knowledge generated during research becomes available to others. In our setting, the territory explored internally and the ideas communicated to the community can be quite different.

[Gans \(2025\)](#) introduces AI into this framework and examines how its capabilities affect research directions. AI can lower research costs within established domains and improve the predictions available to downstream users. These changes can affect researchers’ incentives to pursue novel questions. His analysis already identifies an important distinction between moderately capable AI, which can constrain research ambition, and sufficiently capable AI, which can encourage greater novelty. Our comparison adds a communication margin to this discussion. We ask how society should value faster exploration when only a narrow part of the resulting knowledge is communicated. The capability threshold in our model comes from weighing the deepest usable achievement against the breadth of public knowledge. We also distinguish a multiplicative improvement in search from an improvement in how difficulty increases with depth.

Our representation of ideas on a grid also has precedents in models of technological change. [Silverberg and Verspagen \(2005\)](#) study local search on a lattice of possible technologies, where discoveries become usable through connections to supporting technologies. Their framework helps motivate our picture of research as building connections between neighboring ideas. We use a grid with all nearest-neighbor connections present, introduce increasing discovery difficulty, and examine the social consequences of communicating different portions of the explored region. A related literature studies how knowledge spreads between people. For example, [Fogli and Veldkamp \(2021\)](#) examine how social networks transmit both technologies and disease. Our network instead represents the concepts through which research progresses.

The mathematical foundation of our analysis comes from [Bubeck and Gwynne \(2016\)](#). They study a spatially weighted Eden growth model, represented through first-passage percolation with independent exponential waiting times on edges. The rates vary with location according to a homogeneous function. For rate functions homogeneous of degree below one, their Theorem 1.7 establishes that the rescaled growing region approaches a deterministic shape. We use this theorem together with their construction of the limiting distance and its scaling properties. For rates that decrease with distance as in our model, their framework gives the spatial growth scale $t^{1/(\alpha+1)}$. This growth law is an essential input to our results.

We adapt their theorem to the geometric waiting times generated by repeated research attempts. We then use the limiting shape to evaluate the achievements and public knowledge produced by

exploration. This requires additional arguments for the area of the human community’s map, the length of a shortest established path, and the moment convergence needed to compare expected social values. We prove that a shortest communicated path has length asymptotic to the deepest result, including when the discovered map contains holes and detours. Combining these results with our valuation and two kinds of capability advantage gives the welfare thresholds and their bounds. The growth theorem supplies the geometry. The economic comparison also requires choices about what is communicated and how society values depth and public knowledge.

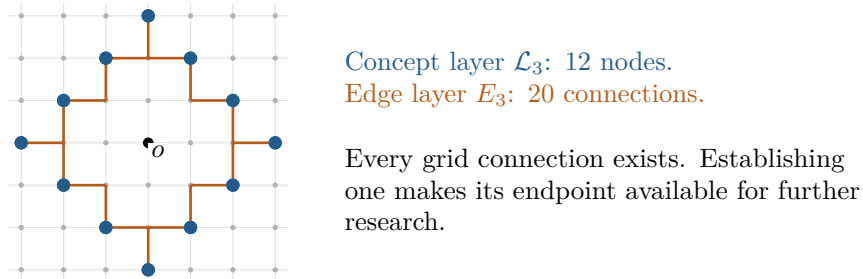
Our motivation also draws directly on the discussion of mathematical understanding. [Thurston \(1994\)](#) describes progress as a collective process sustained by communicating ideas and ways of thinking. [Tao \(2026b\)](#) considers how AI might support explanation, understanding, and the mathematical community. [Gowers \(2026\)](#) emphasizes that mathematicians differ in how they value problem-solving and conceptual understanding. He also raises the possibility that greater AI output could increase the amount of mathematics people assimilate even while leaving more work undigested. These discussions already articulate much of the tension that interests us. Our hope is to give part of that tension a tractable economic form, making explicit how capability, communication, and the value placed on public knowledge jointly determine the comparison.

2 Model

Concepts and connections. The landscape of ideas is the full grid $\mathcal{V} = \mathbb{Z}^2$, with initial knowledge $o = (0, 0)$. The edge set $\mathcal{E}$ joins all horizontal and vertical nearest neighbors. A node represents a concept, technique, or result. Every candidate connection exists, but a researcher must establish it before using it to reach a new concept. Thus an unsuccessful attempt can be revisited. Define

$$d(v) = |v_1| + |v_2|, \quad r(e) = \max\{d(u), d(v)\} \quad \text{for } e = \{u, v\}.$$

The concept layer $\mathcal{L}_n = \{v : d(v) = n\}$ contains $4n$ vertices for $n \geq 1$. The edge layer $E_n = \{e : r(e) = n\}$ contains $8n - 4$ edges. These layers record distance from initial knowledge. Exploration can proceed through any established connection, leaving branches and holes.



The human success probability and the AI success probability are

$$p_H(e) = (1 + r(e))^{-\alpha}, \quad p_A(e) = \min\{1, A(1 + r(e))^{-\alpha/a}\}, \quad \alpha > 0, \quad A, a \geq 1. \quad (2)$$

Attempts are independent across edges and periods within each mode. Both modes retain their discoveries and explore from every known concept. No layer needs to be completed before another can be investigated.

The parameter A gives a *multiplicative advantage*. When $a = 1$, AI succeeds A times as often on a difficult connection, subject to the probability cap. The parameter a gives a *difficulty-exponent advantage*. At $a > 1$, success probabilities decline more slowly with depth. Beyond the capped region,

$$\frac{p_A(e)}{p_H(e)} = A(1 + r(e))^{\alpha(1-1/a)}. \quad (3)$$

AI's relative advantage then increases as research moves farther from the origin. This is the sense in which changing the exponent differs from raising a multiplier. Both success laws are power laws in depth. We hold A and a fixed and finite in all long-run comparisons.

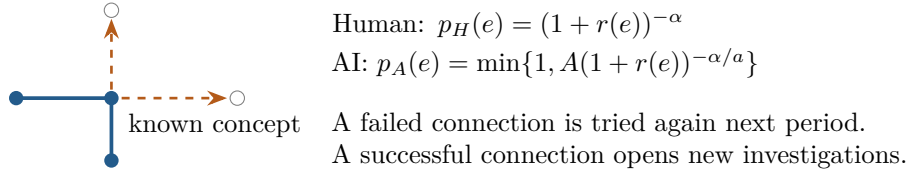


Figure 2: Independent attempts proceed in parallel from established concepts.

Discovery and communication. Let $V_i(t)$ and $F_i(t)$ be the reached vertices and established edges under mode $i \in \{H, A\}$. Its deepest discovery is

$$D_i(t) = \max_{v \in V_i(t)} d(v).$$

The human community communicates its full repertoire, $P_H(t) = V_H(t) \setminus \{o\}$. AI communicates the nonorigin vertices of a shortest path, measured in number of edges in $F_A(t)$, from o to a deepest reached vertex. Fix deterministic rules for choosing among tied targets and paths. The path may turn back toward the origin or take detours. Its edges must have been established by time t .

A communicated concept is assumed to be explained well enough for others to use. Papers are one channel, alongside presentations, workshops, seminars, and informal discussion. A private concept has no public-knowledge value until it is communicated. At each assessment date we evaluate one terminal AI report. Different dates represent alternative times to release that report. An archive of repeated reports would create a union of public concepts and requires a different communication rule.

Social value. For a deepest discovery D and public repertoire P , let

$$U(D, P) = \lambda D^\beta + \mu |P|, \quad \beta \geq 1, \quad \lambda, \mu > 0. \quad (4)$$

The first term values the best usable outcome. Linear value corresponds to $\beta = 1$, and $\beta > 1$ gives convex rewards for more distant breakthroughs. We use maximum depth to summarize direct

achievement in both modes. The second term credits every distinct communicated concept once, including concepts encountered on unsuccessful approaches and on the successful path.

We interpret μ as the expected additional benefit of making one concept available for other questions. For example, each concept may support a separate future use with a common expected value. Additivity then gives $\mu|P|$ by linearity of expectation. Independence of those future uses is unnecessary. This specification assumes a common expected contribution per concept and abstracts from substitution and complementarity between ideas. It gives an extensive measure of the public knowledge stock. An extra communicated concept adds μ to value, and explaining the same concept again adds no further count.

The usable breakthrough and these additional uses are distinct benefits. Giving a concept public-knowledge credit does not add another copy of the original breakthrough payoff. At the same time, this is a substantive valuation assumption. If the benefits of communicating ideas are already included in λD^β , they should be removed from that term before the two contributions are added. We treat research and communication costs as outside the present comparison.

Write $U_i(t) = U(D_i(t), P_i(t))$ and define

$$R(t, A, a) = \frac{\mathbb{E}[U_A(t, A, a)]}{\mathbb{E}[U_H(t)]}, \quad t \geq 1. \quad (5)$$

This is a ratio of expected social values. The denominator is positive for every $t \geq 1$. The model follows the growth of the research stock over time. It values its potential further uses through μ , while later researchers' choices remain outside the exploration process.

3 Main results

The comparison rests on how far research reaches and how much of the explored region becomes public knowledge. We first state the growth result that connects those quantities, then give the quadratic benchmark and the complete characterization for $\beta \geq 1$. Proofs appear in Appendix A.

For a generic research process with parameters (b, ξ) , let $p_{b,\xi}(e) = \min\{1, b(1 + r(e))^{-\xi}\}$, where $b \geq 1$ and $\xi > 0$. Let S_ξ denote its deterministic limiting shape at unit multiplier, and write

$$\kappa_\xi = \frac{1}{\xi + 1}, \quad d_\xi = \max_{x \in S_\xi} \|x\|_1, \quad s_\xi = \text{Area}(S_\xi), \quad c_\alpha = \frac{s_\alpha}{d_\alpha^2}. \quad (6)$$

The appendix constructs S_ξ from the time-constant norm of ordinary first-passage percolation. Human research has $(b, \xi) = (1, \alpha)$, while AI has $(b, \xi) = (A, \alpha/a)$.

Proposition 1 (Depth, breadth, and a communicated path). *Let $L_{b,\xi}(t)$ be the number of nonorigin concepts on a shortest established path to a deepest reached vertex. For every fixed $b \geq 1$ and $\xi > 0$,*

$$\frac{D_{b,\xi}(t)}{(bt)^{\kappa_\xi}} \longrightarrow d_\xi, \quad \frac{|V_{b,\xi}(t) \setminus \{o\}|}{(bt)^{2\kappa_\xi}} \longrightarrow s_\xi, \quad \frac{L_{b,\xi}(t)}{(bt)^{\kappa_\xi}} \longrightarrow d_\xi. \quad (7)$$

Each convergence holds almost surely and in L^p for every finite $p \geq 1$.

The public human map has two-dimensional extent. A successful AI path has length asymptotic to its endpoint's depth. This latter assertion allows finite-time detours. The proof uses the fact that continued research fills in connections within the older part of the map, making short routes available there.

Proposition 2 (Bounds on the area coefficient). *The human area coefficient satisfies*

$$2^{1-2/(\alpha+1)} \leq c_\alpha \leq 2. \quad (8)$$

The upper bound is the area $2D^2$ of the full ℓ^1 diamond of radius D . The lower bound comes from a diamond that the limiting research shape must contain. The coefficient depends on the geometry of exploration and can be bounded without knowing its exact shape.

3.1 A quadratic benchmark with a multiplicative advantage

Take $\beta = 2$ and $a = 1$. Both modes face the same increase in difficulty, and AI has the multiplier A . This case puts breakthrough value and human public knowledge on the same growth scale.

Theorem 1 (The multiplicative capability threshold). *Suppose $\beta = 2$ and $a = 1$. For every fixed $A \geq 1$,*

$$\lim_{t \rightarrow \infty} R(t, A, 1) = \frac{\lambda}{\lambda + \mu c_\alpha} A^{2/(\alpha+1)}. \quad (9)$$

Let

$$A^* = \left(1 + \frac{\mu}{\lambda} c_\alpha\right)^{(\alpha+1)/2} > 1. \quad (10)$$

For $A < A^*$, expected human value exceeds expected AI value at every sufficiently large date. For $A > A^*$, the reverse holds. At $A = A^*$, the ratio converges to one.

Human value is asymptotically $(\lambda + \mu c_\alpha) D_H(t)^2$. AI's depth is asymptotically $A^{1/(\alpha+1)} D_H(t)$, so its leading value is $\lambda A^{2/(\alpha+1)} D_H(t)^2$. The communicated AI path still receives credit, but its number of concepts has order $D_A(t)$. These observations give both the ratio and the threshold.

This is the trade-off that first motivated us. A modestly faster AI can produce a deeper result and still generate less social value overall. When μ/λ is large, the human community's broader repertoire requires a greater discovery advantage to compensate for it. When a breakthrough has unusually large direct benefits, A^* falls. The bounds in (8) also bound the threshold:

$$\left(1 + \frac{\mu}{\lambda} 2^{1-2/(\alpha+1)}\right)^{(\alpha+1)/2} \leq A^* \leq \left(1 + \frac{2\mu}{\lambda}\right)^{(\alpha+1)/2}. \quad (11)$$

Equality of limiting ratios at the threshold leaves the sign of finite-date differences open.

3.2 The general comparison

Set

$$h = \frac{1}{\alpha + 1}, \quad \eta = \frac{\alpha}{a}, \quad k = \frac{1}{\eta + 1} = \frac{a}{a + \alpha}, \quad \Delta = \beta k - h \max\{\beta, 2\}. \quad (12)$$

The exponent h describes human depth growth, and k describes AI depth growth. At $\beta = 1$, AI's path and its breakthrough contribute at the same order. For $\beta > 1$, its breakthrough is the leading contribution.

Theorem 2 (Multiplicative and difficulty-exponent advantages). *For every fixed $A, a \geq 1$ and $\beta \geq 1$,*

$$R(t, A, a) \sim \frac{\lambda d_\eta^\beta (At)^{\beta k} + \mu d_\eta (At)^k}{\lambda d_\alpha^\beta t^{\beta h} + \mu s_\alpha t^{2h}}. \quad (13)$$

Consequently, $R(t, A, a) \sim Ct^\Delta$ for a finite positive constant C . A negative Δ gives $R \rightarrow 0$, and a positive Δ gives $R \rightarrow \infty$. More explicitly:

1. If $a = 1$, then

$$\lim_{t \rightarrow \infty} R(t, A, 1) = \begin{cases} 0, & 1 \leq \beta < 2, \\ \frac{\lambda A^{2/(\alpha+1)}}{\lambda + \mu c_\alpha}, & \beta = 2, \\ A^{\beta/(\alpha+1)}, & \beta > 2. \end{cases} \quad (14)$$

2. If $a > 1$ and $\beta \geq 2$, then $R(t, A, a) \rightarrow \infty$.

3. If $1 \leq \beta < 2$ and $\beta(\alpha + 1) \leq 2$, then $R(t, A, a) \rightarrow 0$ for every fixed finite A, a .

4. If $1 \leq \beta < 2$ and $\beta(\alpha + 1) > 2$, define

$$a_c = \frac{2\alpha}{\beta(\alpha + 1) - 2} > 1. \quad (15)$$

Then $R \rightarrow 0$ for $a < a_c$, and $R \rightarrow \infty$ for $a > a_c$, for every fixed finite $A \geq 1$. At $a = a_c$, the ratio converges to the positive constant in Theorem 3.

The same asymptotic formula holds almost surely for the ratio of realized values, once human value becomes positive.

The theorem distinguishes the roles of the two capability parameters. A fixed multiplier changes the leading coefficient. A smaller difficulty exponent changes the growth rate itself. Consequently, social weights can decide the limiting ranking when the growth rates match. Away from those boundaries, the weights still affect value and the timing of a possible crossover, while the time exponents determine the eventual ranking.

For $1 \leq \beta < 2$, human public knowledge grows faster than human breakthrough value. Under a purely multiplicative advantage, AI's deeper result cannot make up for this difference indefinitely. The exponent advantage can change that conclusion. When $\beta(\alpha + 1) > 2$, crossing a_c makes AI breakthrough value grow faster than human public knowledge. A more severe baseline difficulty α makes this kind of improvement particularly consequential.

The condition for a finite a_c also has an intuitive source. Any discovery must traverse successive edges, so $D_A(t) \leq t$. For fixed finite a , AI depth grows at a power strictly below one. When human public knowledge grows at least as fast as t^β , AI cannot overtake it with this terminal-report rule. This is exactly the region $\beta(\alpha + 1) \leq 2$. Letting capability depend on the time horizon would require another analysis.

For $\beta > 2$ and $a = 1$, every $A > 1$ gives a limiting AI advantage. At $A = a = 1$, both modes have the same discovery law. Human communication then adds value at finite dates, while the ratio tends to one because breakthroughs dominate the two totals. The distinction between a ratio of one and finite-date equality is useful here.

3.3 When growth rates match

At the exponent threshold, a simple cancellation gives an explicit constant. Define

$$\theta = \frac{2}{\alpha + 1}, \quad w_\beta = \begin{cases} 1 + \lambda/\mu, & \beta = 1, \\ \lambda/\mu, & 1 < \beta < 2. \end{cases} \quad (16)$$

Theorem 3 (The multiplicative threshold at a_c). *Suppose $1 \leq \beta < 2$ and $\beta(\alpha + 1) > 2$. At $a = a_c$,*

$$\lim_{t \rightarrow \infty} R(t, A, a_c) = \frac{w_\beta}{c_\alpha} \left(\frac{A\beta}{2} \right)^{2/(\alpha+1)}. \quad (17)$$

Define

$$A_c = \frac{2}{\beta} \left(\frac{c_\alpha}{w_\beta} \right)^{(\alpha+1)/2}. \quad (18)$$

The limiting ratio is below one for $A < A_c$, above one for $A > A_c$, and equal to one for $A = A_c$. If $A_c < 1$, all admissible multipliers $A \geq 1$ yield a limiting ratio above one.

For $1 < \beta < 2$, human value is led by public knowledge and AI value by its deepest discovery. A larger μ/λ therefore raises A_c . At $\beta = 1$, both parts of AI's value matter, explaining the term $1 + \lambda/\mu$. Communicating a long path itself creates a substantial repertoire in that case.

The explicit coefficient uses more than the growth exponents. For a common constant $\tau_* > 0$ determined by ordinary first-passage percolation, the deepest radius of the weighted shape is

$$d_\xi = \left(\frac{\xi + 1}{\tau_*} \right)^{1/(\xi+1)}.$$

At a_c , the relation $\eta + 1 = \beta(\alpha + 1)/2$ gives

$$d_\eta^\beta = (\beta/2)^{2/(\alpha+1)} d_\alpha^2.$$

The unknown fastest-direction constant cancels. The remaining geometric quantity is the human area coefficient c_α , whose bounds are given in Proposition 2.

4 Simulations

The figures follow the same retry process as the model. We assign an independent unit-exponential mark E_e to every edge as it is first encountered and convert it to a geometric duration $G_e = \lceil E_e / [-\log(1 - p_e)] \rceil$, taking $G_e = 1$ when $p_e = 1$. Dijkstra's algorithm computes arrival times. An edge is established at $\min\{T(u), T(v)\} + G_e$, so the recorded graph includes connections completed after both endpoints are known. Breadth-first search in that graph finds the shortest communicated path to a deepest node.

We simulate the infinite lattice by generating edges as needed. The calculation has no spatial truncation. Within each replication the research modes use common independent edge marks, and different replications use independent seeds. The reproduction files retain the parameters, seeds, raw outcomes, and plotted summaries. Twenty-one small examples agree exactly with explicit period-by-period updates for both arrival times and established edges. A probability-one check recovers the deterministic grid diamond.

Figure 3 shows one realization. The human map remains the same across panels, while the AI parameters change. Pale points show the much broader region that AI explored privately. Its communicated path is a small part of that region. The displayed depths and node counts determine the two welfare terms directly. For example, the human sample has $D_H = 25$ and $|P_H| = 781$. With $\lambda = \mu = 1$ and $\beta = 2$, its value is $625 + 781$. At $(A, a) = (3, 1)$, AI has depth 46 and communicates 46 concepts, giving value $2116 + 46$.

The same research landscape, different capability and communication

Exact simulation at $t = 100$ | $\alpha = 1, \beta = 2, \lambda = \mu = 1$

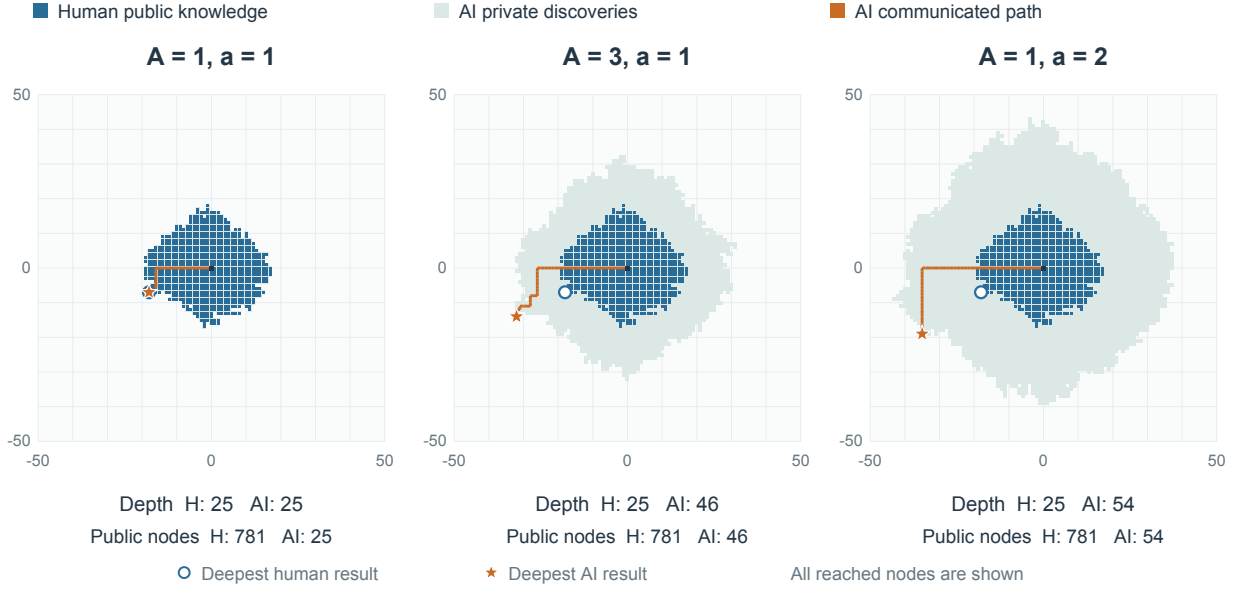


Figure 3: One simulated landscape at $t = 100$, with $\alpha = 1, \beta = 2$, and $\lambda = \mu = 1$. Blue concepts are communicated by humans. Pale concepts show AI's private discoveries, and orange gives its shortest established path to a deepest result. All nodes share the same independent edge marks. Exploration runs on the infinite lattice, and all reached concepts are displayed.

Figure 4 reports ratios of mean utilities over 48 independent paired replications. It illustrates how a larger multiplier can overcome the communication gap in the quadratic benchmark and how reducing the difficulty exponent produces a growing advantage over the displayed horizons. The uncertainty bars describe simulation error in these finite-date estimates. The theorems establish the limiting comparisons. Figure 5 plots those analytical regimes separately, including the region where no fixed finite exponent advantage can overturn human breadth.

Multiplicative advantages and changes in the difficulty exponent

$\alpha = 1, \beta = 2, \lambda = \mu = 1$ | 48 independent coupled replications

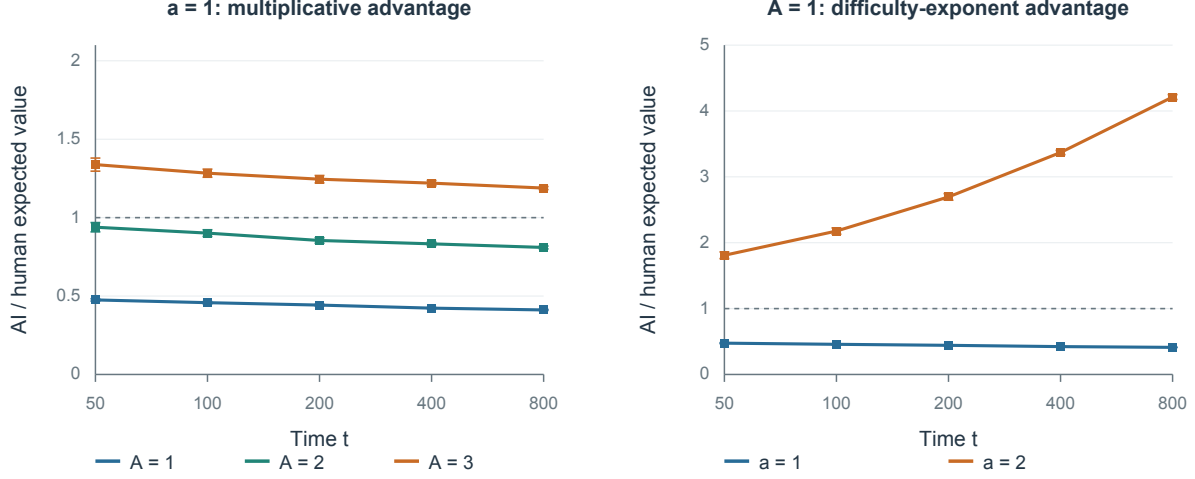


Figure 4: Simulated AI-to-human value ratios for $\alpha = 1, \beta = 2$, and $\lambda = \mu = 1$. Points are ratios of mean utilities across 48 independent coupled replications. Bars give approximate 95% paired delta-method confidence intervals. The dashed line indicates equal value. Dates are spaced logarithmically. These are finite-time simulation results.

Long-run regimes with $\beta \geq 1$

Analytical phase diagram | every fixed finite $A \geq 1$ and $\lambda, \mu > 0$

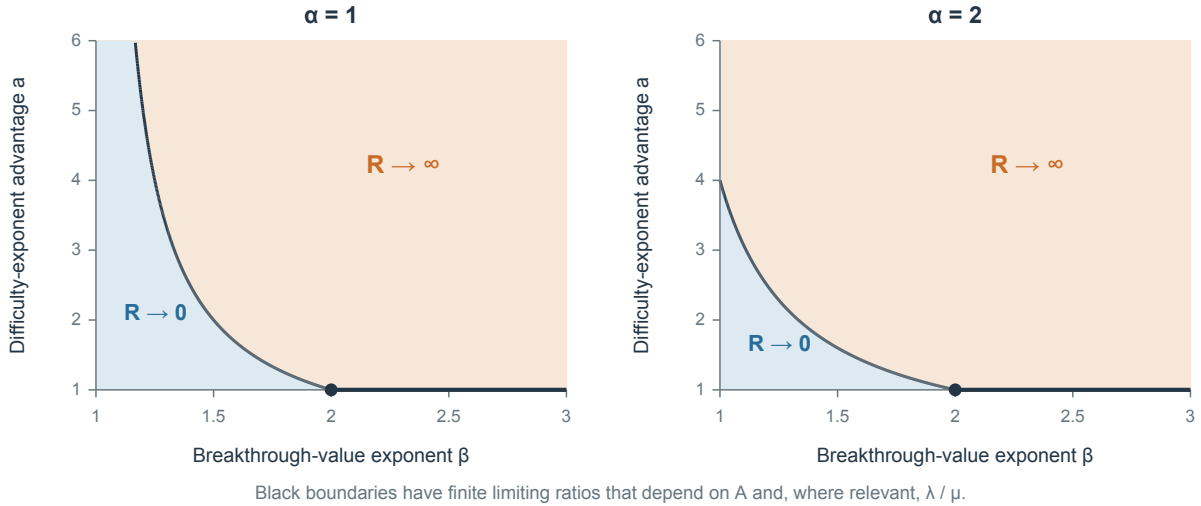


Figure 5: Analytical long-run regimes for $\beta \geq 1$. Blue and orange regions have limits zero and infinity, respectively, for every fixed finite $A \geq 1$. The separating curve and the boundary $a = 1, \beta \geq 2$, have finite limiting ratios. Their ranking depends on the coefficients characterized in the theorems.

5 Discussion and conclusion

We began with a concern about what society gains when difficult problems are solved faster than the knowledge surrounding them is communicated. The model gives this concern a simple economic form. In the quadratic benchmark, a modest discovery advantage leaves human research more valuable when it communicates a much broader map. A sufficiently large multiplier compensates through a deeper breakthrough. The threshold depends on what a community values, and those values can differ even within a discipline. A mathematician hoping to develop methods for other questions and a patient waiting for an effective treatment may reasonably place different weights on the benefits research produces.

The general model asks a further question about what it means for AI to become more capable. A multiplicative improvement accelerates research at every depth. An improvement in the difficulty exponent makes AI’s relative advantage increase as the problems become more remote. These changes have different long-run consequences. With quadratic breakthrough value, any fixed exponent advantage eventually outweighs the communication gap. When breakthrough value grows less than quadratically, that conclusion requires a sufficiently large exponent advantage and sufficiently strong baseline difficulty. With more than quadratic value, even a fixed multiplier above one yields a limiting advantage. Our hope is that these distinctions make discussions of “more capable AI” a little more precise.

We expect actual research to combine elements of both benchmarks. People can use AI to reach further, investigate the surrounding ideas, and communicate what they learn. The model supplies a simple reason to value that effort.

Proposition 3 (Additional communication). *At a fixed discovery state with depth D , if $P \subseteq P'$, then*

$$U(D, P') - U(D, P) = \mu |P' \setminus P|.$$

This follows directly from (4). It is a statement about the benefit of additional usable knowledge. A recommendation about how much to communicate would also need the time and effort required to explain ideas and any consequences for discovery.

Our single-report assumption deliberately stacks the deck against AI along the communication dimension. Human researchers share their whole map, while AI shares one successful path. In practice, repeated reports could communicate a much broader union of ideas. Holding discoveries fixed, expanding that public repertoire increases AI’s value and can lower the capability needed to exceed the human benchmark. What interests us is that even this demanding comparison can favor AI when its discoveries are sufficiently deep. We assess direct achievement by the deepest result under both research modes, so this conclusion does not require crediting AI with a larger number of separately valued breakthroughs.

Several choices keep the model deliberately spare. A node count gives every communicated concept the same expected additional value. Concepts can differ greatly in usefulness, and their benefits may overlap or depend on being understood together. Maximum depth is also an imperfect summary of the practical value of research. The grid gives knowledge two dimensions, and parallel attempts allow research effort to expand with the explored map. Costs, verification, researcher incentives, and the feedback from public knowledge into subsequent research choices would all matter in a fuller account. The thresholds describe the consequences of these primitives. Applying them to a particular field would require evidence about both capability and value.

Perhaps people on different sides of this debate can find some solace in the comparison. Mathematicians troubled by the seemingly pernicious effects of unfettered AI research may believe that

public knowledge deserves a much larger weight than a completed answer and that present systems have yet to compensate for what they leave uncommunicated. The quadratic benchmark makes that position intelligible. Those more optimistic about AI may see scope for discovery to become so much faster and deeper that even a narrow account of the process creates enormous value. Our results also make room for that possibility, while showing which kinds of capability improvement support it. We hope the model helps us think more carefully about both what we want AI to discover and what we want the people who come after us to understand.

A Proofs

We establish the growth of the discovered map and the length of a shortest communicated path before comparing their social values. The distinction between these two objects matters at $\beta = 1$, when AI's communicated path and its deepest achievement have the same order of value.

Throughout the first three subsections, fix a difficulty exponent $\xi > 0$ and a multiplier $b \geq 1$. The success probability is $p_e = \min\{1, b(1 + r(e))^{-\xi}\}$. Write

$$\kappa = \frac{1}{\xi + 1}, \quad \delta = \frac{2}{\xi + 2}, \quad R_t = (bt)^\kappa.$$

The quantities $D(t)$, $V(t)$, and $L(t)$ denote maximum attained depth, the set of reached vertices, and the number of edges in a shortest established path to a deepest vertex. Thus $L(t)$ is also the number of nonorigin vertices communicated along that path. All constants below may depend on the fixed parameters.

A.1 Passage times and limiting geometry

Assign independent $E_e \sim \text{Exp}(1)$ to the lattice edges. Define

$$X_e = \frac{(1 + r(e))^\xi}{b} E_e, \quad G_e = \begin{cases} \lceil E_e / [-\log(1 - p_e)] \rceil, & p_e < 1, \\ 1, & p_e = 1. \end{cases} \quad (19)$$

Then G_e is geometric on $\{1, 2, \dots\}$ with success probability p_e . Its clock starts when the first endpoint of e is reached. For positive edge weights $Z = (Z_e)$, put

$$T_Z(v) = \inf_{\pi: o \rightarrow v} \sum_{e \in \pi} Z_e, \quad V_Z(t) = \{v : T_Z(v) \leq t\}.$$

First-arrival times in the research process are exactly $T_G(v)$. Indeed, an edge completes at time $\min\{T_G(u), T_G(v)\} + G_e$, which is the same recursion as independent repeated attempts from its first reached endpoint. Continued attempts when both endpoints are known add connections without altering these first-arrival times. Every prefix of a minimizing G -path is minimizing, so every edge of such a path is established by the arrival time at its terminal vertex. Since $G_e \geq 1$, minimizing G -paths exist and

$$D(t) \leq L(t) \leq t, \quad |V(t)| \leq 1 + 2t(t + 1) \quad (t \in \mathbb{N}). \quad (20)$$

Before the first discovery we set $D(t) = L(t) = 0$.

Lemma 1 (Length of a fast exponential path). *Let $M_X(t)$ be the largest number of edges in a simple path from o whose total X -time is at most t . There are constants $K, c, C > 0$ such that, for all sufficiently large t ,*

$$\mathbb{P}(M_X(t) \geq Kt^\delta) \leq C \exp(-ct^\delta).$$

In particular, finite X -passage times are attained by finite simple paths, and $M_X(t) = o(t)$ almost surely.

Proof. There are at most $C_0 r^2$ lattice edges with $r(e) < r$. Choose $c_0 > 0$ so that fewer than $m/2$ edges have $r(e) < c_0 \sqrt{m}$ for all large integers m . Every simple m -edge path uses at least $j = \lceil m/2 \rceil$ distinct edges whose exponential X -rates are at most $C_1 m^{-\xi/2}$. The sum of these times stochastically dominates $m^{\xi/2} Z / C_1$, where Z is gamma with shape j and unit rate. The density bound and $j! \geq (j/e)^j$ give

$$\mathbb{P}(Z \leq y) \leq \frac{y^j}{j!} \leq \left(\frac{ey}{j}\right)^j.$$

Consequently, for a fixed simple path of length m , its probability of having X -time at most t is bounded by $[C_2 t / m^{1+\xi/2}]^{m/2}$ whenever this bracket is at most one. There are at most $4 \cdot 3^{m-1}$ such paths from o . Since $\delta(1 + \xi/2) = 1$, a union bound followed by a sum over $m \geq Kt^\delta$ gives the asserted estimate when K is large enough. The bound is summable over integer times, so the Borel–Cantelli lemma and monotonicity give the almost-sure conclusion.

For fixed t , the same calculation tends to zero as $m \rightarrow \infty$. Almost surely there are therefore no arbitrarily long simple paths of bounded X -time. Local finiteness of the lattice then ensures that the infimum defining a finite X -passage time is attained. $\square$

Let ν be the time-constant norm of ordinary nearest-neighbor first-passage percolation with independent unit-rate exponential edge times. Define

$$\rho_\xi(x, y) = \inf_{\gamma: x \rightarrow y} \int_0^1 \|\gamma(s)\|_1^\xi \nu(\gamma'(s)) ds, \quad S_\xi = \{x : \rho_\xi(o, x) \leq 1\}, \quad (21)$$

where the infimum is over piecewise linear curves. Also set

$$d_\xi = \max_{x \in S_\xi} \|x\|_1, \quad s_\xi = \text{Area}(S_\xi), \quad \tau_* = \min_{\|u\|_1=1} \nu(u). \quad (22)$$

Lemma 2 (Geometry and area bounds). *The set S_ξ is compact and star-shaped, contains a neighborhood of the origin, and has a continuous positive radial function. Its boundary has area zero. Moreover,*

$$d_\xi = \left(\frac{\xi + 1}{\tau_*}\right)^{1/(\xi+1)}. \quad (23)$$

For $c_\xi = s_\xi / d_\xi^2$,

$$2^{1-2/(\xi+1)} \leq c_\xi \leq 2. \quad (24)$$

Proof. Spatial rescaling in (21) gives $\rho_\xi(o, cx) = c^{\xi+1} \rho_\xi(o, x)$ for $c > 0$. Norm equivalence bounds this quantity above and below by positive constant multiples of $\|x\|_1^{\xi+1}$. The upper bound follows from the straight segment, and the lower bound follows by integrating the change in radial distance along any admissible curve. The weighted distance is continuous. Away from the origin this follows by appending a short line segment on a bounded region, and the upper bound proves continuity at the origin. These facts give compactness, a continuous positive radial function, and star-shapedness. Every ray meets the boundary once, so polar integration shows that the boundary has area zero.

For the exact depth constant, $\nu(z) \geq \tau_* \|z\|_1$ for every z . If $r(s) = \|\gamma(s)\|_1$, then $|r'(s)| \leq \|\gamma'(s)\|_1$ almost everywhere. A curve from o to depth r consequently has weighted length at least

$$\tau_* \int_0^1 r(s)^\xi |r'(s)| ds \geq \frac{\tau_* r^{\xi+1}}{\xi + 1}.$$

Choose a direction u_* attaining the minimum in (22). The straight segment to ru_* attains this lower bound. This proves (23).

Write $c_0 = \nu((1, 0)) = \nu((0, 1))$. The symmetries and convexity of ν imply

$$\frac{c_0}{2} \|z\|_1 \leq c_0 \|z\|_\infty \leq \nu(z) \leq c_0 \|z\|_1.$$

For example, reflection in the second coordinate and convexity show $\nu((z_1, 0)) \leq \nu(z)$, and the other coordinate is analogous. Thus $c_0/2 \leq \tau_* \leq c_0$. Straight segments show that S_ξ contains the ℓ^1 -ball of radius $r_- = [(\xi + 1)/c_0]^{1/(\xi+1)}$, and it lies in the ℓ^1 -ball of radius d_ξ . Such a ball has area twice its squared radius. Hence

$$2(\tau_*/c_0)^{2/(\xi+1)} \leq \frac{s_\xi}{d_\xi^2} \leq 2,$$

which implies (24). □

A.2 The shape theorem and moment convergence

The quantitative form of the weighted shape theorem is useful here. It controls the rare realizations in which a research map or a shortest path is unusually large.

Lemma 3 (Shape and moments of the research map). *For every $\varepsilon \in (0, 1)$ and $N > 0$, there is a finite $C_{\varepsilon, N}$ such that, for every sufficiently large T ,*

$$\mathbb{P} \left(\begin{array}{c} (1 - \varepsilon)R_t S_\xi \cap \mathbb{Z}^2 \subseteq V(t) \subseteq (1 + \varepsilon)R_t S_\xi \cap \mathbb{Z}^2 \\ \text{for every integer } t \geq T \end{array} \right) \geq 1 - C_{\varepsilon, N} T^{-N}. \quad (25)$$

Consequently,

$$\frac{D(t)}{R_t} \longrightarrow d_\xi, \quad \frac{|V(t) \setminus \{o\}|}{R_t^2} \longrightarrow s_\xi \quad (26)$$

almost surely and in L^p for every finite $p \geq 1$. In particular, for every $q > 0$, $\mathbb{E}[D(t)^q]/R_t^q \rightarrow d_\xi^q$.

Proof. For the midpoint m_e of edge e , define

$$Y_e = \frac{\|m_e\|_1^\xi}{b} E_e.$$

Theorem 1.7 of [Bubeck and Gwynne \(2016\)](#) applies to the rate function $f(x) = \|x\|_1^{-\xi}$. This function is positive, locally Lipschitz away from zero, and homogeneous of degree $-\xi < 1$. Its limiting shape is (21). The theorem gives relative shape error of order $t^{-\chi}$ for any $0 < \chi < 1/[3(\xi + 1)]$, simultaneously for all $t \geq T$, outside an event whose probability decays faster than every negative power of T . The constant multiplier b is absorbed by the time change $t \mapsto bt$. Their convention of attaching unit boxes to reached vertices gives the same lattice inclusions for any fixed positive relative error once T is sufficiently large.

We explain why these probability bounds also hold for the geometric process. Since $\|m_e\|_1 = r(e) - 1/2$, we have $Y_e \leq X_e$, and, for every fixed $\zeta > 0$, $X_e \leq (1 + \zeta)Y_e$ outside a finite set F_ζ . Set $C_\zeta = \sum_{e \in F_\zeta} X_e$. Evaluating simple minimizing paths gives

$$T_Y(v) \leq T_X(v) \leq (1 + \zeta)T_Y(v) + C_\zeta. \quad (27)$$

The required minimizing paths exist by Lemma 1, since $Y_e \geq 4^{-\xi} X_e$. The finite sum C_ζ has an exponential tail.

The coupling (19) gives $G_e \leq X_e + 1$ globally. For all sufficiently distant edges, $p_e \leq \zeta < 1$ and

$$-\log(1 - p_e) \leq \frac{p_e}{1 - p_e}, \quad G_e \geq (1 - \zeta)X_e.$$

The finitely many exceptions can again be absorbed into a finite sum C'_ζ of exponential variables. Thus

$$T_G(v) \geq (1 - \zeta)T_X(v) - C'_\zeta, \quad V_X(t) \subseteq V_G(t + M_X(t)). \quad (28)$$

In the capped case $p_e = 1$, the inequality $G_e \leq X_e + 1$ remains valid, and these edges belong to the finite exceptional set.

For fixed ζ , the probability that either exceptional sum exceeds ζT is at most $C_\zeta e^{-c_\zeta T}$. By Lemma 1 and a union bound over integer times, the probability that $M_X(s) > \zeta s$ for some integer $s \geq T$ decays faster than every negative power of T . Monotonicity handles the bounded rounding of real times in (28). It follows that, outside an event of this same small probability, the two comparisons sandwich $V_G(t)$ between exponential clusters at times $(1 - O(\zeta))t$ and $(1 + O(\zeta))t$, simultaneously for every integer $t \geq T$. First choose ζ small enough for the desired ε , then apply the quantitative exponential shape theorem. This proves (25).

Taking a countable sequence $\varepsilon \downarrow 0$ gives the almost-sure shape inclusions. Maximum depth converges by compactness and lattice approximation. Since ∂S_ξ has area zero, lattice-point Riemann sums between the two dilations give the area limit in (26).

For moment convergence, fix a positive tolerance in either limit. The shape inclusions and the deterministic Riemann-sum limits imply that the probability of exceeding this tolerance is at most $C_N t^{-N}$ for every N , once t is sufficiently large. On its complement the error is bounded by that tolerance. On the exceptional event, (20) bounds the normalized depth by t/R_t and the normalized area by $[1 + 2t(t + 1)]/R_t^2$. These bounds grow polynomially in t . Choose N larger than the relevant polynomial degree after taking any prescribed moment. The exceptional contribution then tends to zero. Letting the tolerance vanish proves convergence in every finite L^p . The same argument applied to $(D(t)/R_t)^q$ proves the last assertion. $\square$

A.3 The length of a communicated path

An early discovered interior region has time to establish its remaining connections. A short route through this region can be joined to the last part of the discovery route near the frontier. The next lemma makes this argument precise and controls the length of that last part.

Lemma 4 (A shortest established path). *For the shortest-path communication rule,*

$$\frac{L(t)}{R_t} \longrightarrow d_\xi \quad \text{almost surely and in } L^p \quad \text{for every finite } p \geq 1. \quad (29)$$

In particular, $L(t)/D(t) \rightarrow 1$ almost surely and in every finite L^p , with this ratio defined to be one before the first discovery. For each $\varepsilon > 0$ and $N > 0$, the probability that $|L(t)/R_t - d_\xi| > \varepsilon$ is at most $C_{\varepsilon,N} t^{-N}$ for all sufficiently large t .

Proof. Fix $0 < \varepsilon < 1/16$, abbreviate $R = R_t$, and define

$$K_j = (1 - j\varepsilon)RS_\xi \quad (j = 1, 2, 3), \quad t_0 = (1 - \varepsilon/2)^{\xi+1}t, \quad t_- = (1 - 4\varepsilon)^{\xi+1}t.$$

Rounding t_0, t_- to integer times has no effect on the following estimates. The shape lemma, with a sufficiently small fixed relative error, implies all three of the following outside an event of probability at most $C_{\varepsilon,N} t^{-N}$ for every N :

1. Every lattice vertex in K_1 is reached by t_0 .
2. $V(t_-) \subseteq K_3$, and $V(t) \subseteq (1 + \varepsilon)RS_\xi$.
3. $D(t) > (1 - 2\varepsilon)d_\xi R$.

These follow by comparing the scale factors. For example, relative error $\varepsilon/4$ at t_0 gives an inner factor $(1 - \varepsilon/4)(1 - \varepsilon/2) > 1 - \varepsilon$.

Completion of the interior. There are $O(R^2)$ edges with both endpoints in K_1 . Their probabilities satisfy $p_e \geq cR^{-\xi}$ for a fixed $c > 0$. For each such edge,

$$\mathbb{P}(G_e > t - t_0 - 2) \leq \exp(-c_\varepsilon R),$$

because $t - t_0$ is a positive multiple of εt and $t/R^\xi = R/b$. A union bound shows that, outside an event of probability $CR^2e^{-c_\varepsilon R}$, every edge with both endpoints in K_1 is established by time t . This reasoning uses continued attempts when both endpoints have already been reached. No independence between the early-arrival event and these edge durations is required for the union bound.

Length of the final part of a discovery route. Choose a deepest reached vertex v and a minimizing G -path from o to v . The third shape property places v outside K_2 . Let u be the last vertex of this path in K_2 . It is adjacent to a vertex outside K_2 . By the continuous positive radial function of S_ξ , K_3 has distance of order R from the complement of K_2 for fixed ε . Thus $u \notin K_3$ for sufficiently large t . The second shape property gives $T_G(u) > t_-$. Since prefixes of the minimizing path are minimizing, its segment from u to v has total G -time at most

$$t - t_- \leq C_0 \varepsilon t,$$

where C_0 can be chosen independently of $\varepsilon \in (0, 1/16)$. Every edge of this segment has radial depth at least $c_0 R$, with $c_0 > 0$ independent of this range of ε , because S_ξ contains a neighborhood of the origin. Its geometric success probabilities are therefore at most $q_R = C_1 R^{-\xi} < 1$.

For a geometric variable G with success probability $p \leq q_R$, take $\vartheta = 12q_R$. Its Laplace transform satisfies

$$\mathbb{E}[e^{-\vartheta G}] = \frac{p}{e^\vartheta - 1 + p} \leq \frac{1}{12}.$$

For any fixed simple m -edge path using only these exterior edges,

$$\mathbb{P}\left(\sum_{e \in \pi} G_e \leq C_0 \varepsilon t\right) \leq \exp(C_2 \varepsilon R) 12^{-m}.$$

There are $O(R^2)$ possible starting vertices in $(1 + \varepsilon)RS_\xi$, and at most $4 \cdot 3^{m-1}$ simple paths of length m from each one. A union bound over $m \geq K\varepsilon R$, with K sufficiently large and independent of ε , shows that a segment of the type just described has at most $K\varepsilon R$ edges outside an event of probability $CR^2e^{-c\varepsilon R}$. Taking integer ceilings only adds a bounded number of edges.

Joining the two routes. There is a nearest-neighbor path from o to u with exactly $\|u\|_1$ edges that remains within distance two of the line segment $[o, u]$. To see this explicitly, write $m = |u_1| + |u_2|$ and take, for $j = 0, \dots, m$, the vertex with signed coordinates $\lfloor j|u_1|/m \rfloor$ and $j - \lfloor j|u_1|/m \rfloor$. Successive vertices differ by one grid step, and each lies within ℓ^1 -distance two of $(j/m)u$. The segment belongs to K_2 , because S_ξ is star-shaped. For fixed ε and sufficiently large R , the distance-two neighborhood of K_2 lies in K_1 . Every edge of this digital straight path is consequently established by t . Append

the final segment of the discovery route and delete any loops. The resulting established path to v has at most

$$\|u\|_1 + K\varepsilon R + O(1) \leq D(t) + K\varepsilon R + O(1)$$

edges. Together with the deterministic lower bound $L(t) \geq D(t)$, this proves, for every fixed ε and N ,

$$\mathbb{P}(L(t) - D(t) > K\varepsilon R_t + O(1)) \leq C_{\varepsilon, N} t^{-N}. \quad (30)$$

The exponentially small terms above satisfy this bound as well.

Apply Borel–Cantelli for each member of a countable sequence $\varepsilon \downarrow 0$, then use the depth limit to obtain (29) almost surely. The shape concentration and (30) also give the stated probability bound for any prescribed error in $L(t)/R_t$. Finally, $L(t) \leq t$ by (20). Splitting any moment according to that error event, choosing N larger than the resulting polynomial power, and letting the error tolerance vanish proves convergence in every finite L^p . Division by the positive almost-sure depth limit gives $L(t)/D(t) \rightarrow 1$ almost surely. The depth concentration ensures $D(t) \geq d_\xi R_t/2$ outside an event with probability smaller than every negative power of t . On that event’s complement, the normalized path error controls the ratio error. Globally, the ratio is at most $\max\{1, t\}$. The same moment-splitting argument therefore gives its convergence in every finite L^p . $\square$

A.4 The welfare comparison

Proof of Propositions 1 and 2. Lemma 2 establishes the shape constants, the formula for d_ξ , and the bounds on c_α . Lemma 3 establishes the depth and area limits. Apply these results with $(\xi, b) = (\alpha, 1)$ for humans and $(\xi, b) = (\eta, A)$ for AI, where $\eta = \alpha/a$. Lemma 4 gives the communicated path limit. These statements hold simultaneously on a probability-one event under any coupling of the two processes with their specified marginal laws. $\square$

Proof of Theorem 2. Write

$$h = \frac{1}{\alpha + 1}, \quad k = \frac{1}{\eta + 1} = \frac{a}{a + \alpha}.$$

Lemma 3 gives $\mathbb{E}[D_H(t)^\beta] \sim d_\alpha^\beta t^{\beta h}$ and $\mathbb{E}[|P_H(t)|] \sim s_\alpha t^{2h}$. Its AI counterpart gives $\mathbb{E}[D_A(t)^\beta] \sim d_\eta^\beta (At)^{\beta k}$, and Lemma 4 gives $\mathbb{E}[|P_A(t)|] \sim d_\eta (At)^k$. Hence

$$\mathbb{E}[U_H(t)] \sim \lambda d_\alpha^\beta t^{\beta h} + \mu s_\alpha t^{2h}, \quad (31)$$

$$\mathbb{E}[U_A(t)] \sim \lambda d_\eta^\beta (At)^{\beta k} + \mu d_\eta (At)^k. \quad (32)$$

All four coefficients are positive, so adding the two contributions preserves these equivalents even when their powers differ. This proves the stated equivalent for the ratio of expected values. The almost-sure versions of the lemmas give the same equivalent for $U_A(t)/U_H(t)$ once the denominator is positive.

To make the leading coefficient explicit, define

$$H_\beta = \begin{cases} \mu s_\alpha, & 1 \leq \beta < 2, \\ \lambda d_\alpha^2 + \mu s_\alpha, & \beta = 2, \\ \lambda d_\alpha^\beta, & \beta > 2, \end{cases} \quad G_\beta = \begin{cases} (\lambda + \mu) d_\eta, & \beta = 1, \\ \lambda d_\eta^\beta, & \beta > 1. \end{cases}$$

Keeping the largest power in each sum yields

$$R(t, A, a) \sim \frac{G_\beta}{H_\beta} A^{\beta k} t^\Delta, \quad \Delta = \frac{a\beta}{a + \alpha} - \frac{\max\{\beta, 2\}}{\alpha + 1}. \quad (33)$$

At $\beta = 1$ both AI contributions have the same power, and at $\beta = 2$ both human contributions do. Since $G_\beta/H_\beta > 0$, the ratio tends to zero for $\Delta < 0$ and to infinity for $\Delta > 0$.

When $a = 1$, $\eta = \alpha$ and $k = h$. Substitution gives

$$\lim_{t \rightarrow \infty} R(t, A, 1) = \begin{cases} 0, & 1 \leq \beta < 2, \\ \frac{\lambda}{\lambda + \mu c_\alpha} A^{2/(\alpha+1)}, & \beta = 2, \\ A^{\beta/(\alpha+1)}, & \beta > 2. \end{cases}$$

If $a > 1$ and $\beta \geq 2$, then $k > h$ and $\Delta = \beta(k - h) > 0$, proving divergence to infinity.

For $1 \leq \beta < 2$, the sign of Δ is the sign of

$$a[\beta(\alpha + 1) - 2] - 2\alpha.$$

It is negative for every finite a when $\beta(\alpha + 1) \leq 2$. Otherwise its unique zero is

$$a_c = \frac{2\alpha}{\beta(\alpha + 1) - 2}.$$

This zero exceeds one because $2\alpha - [\beta(\alpha + 1) - 2] = (2 - \beta)(\alpha + 1) > 0$. This proves the cases on either side of a_c . Whenever the limiting ratio is strictly below or above one, the corresponding welfare comparison holds at every sufficiently large date. For realized ratios that date may depend on the realization. $\square$

Proof of Theorem 1. Set $\beta = 2$ and $a = 1$ in (31) and (32). Since $s_\alpha = c_\alpha d_\alpha^2$, division gives

$$\lim_{t \rightarrow \infty} R(t, A, 1) = \frac{\lambda}{\lambda + \mu c_\alpha} A^{2/(\alpha+1)}.$$

This expression is strictly increasing in A and equals one at

$$A^* = \left(1 + \frac{\mu}{\lambda} c_\alpha\right)^{(\alpha+1)/2}.$$

The strict limiting comparisons follow immediately. Lemma 2 provides the bounds on c_α and hence the stated bounds on A^* . The same formula shows that the threshold increases with μ/λ for fixed α and exploration geometry. At $A = A^*$ the ratio converges to one. This limit alone does not settle the sign of finite-date differences. $\square$

Proof of Theorem 3. Suppose $1 \leq \beta < 2$ and $\beta(\alpha + 1) > 2$. At $a = a_c$, put $\theta = 2/(\alpha + 1)$. Then $\beta k = \theta$ and $\eta + 1 = \beta(\alpha + 1)/2$. The exact depth constant in Lemma 2 gives

$$d_\eta^\beta = \left(\frac{\beta}{2}\right)^\theta d_\alpha^2. \quad (34)$$

Using $s_\alpha = c_\alpha d_\alpha^2$ in (33) therefore yields

$$\lim_{t \rightarrow \infty} R(t, A, a_c) = \begin{cases} \frac{1 + \lambda/\mu}{c_\alpha} (A/2)^\theta, & \beta = 1, \\ \frac{\lambda/\mu}{c_\alpha} (A\beta/2)^\theta, & 1 < \beta < 2. \end{cases}$$

In the notation of the theorem, use $w_1 = 1 + \lambda/\mu$ and $w_\beta = \lambda/\mu$ for $\beta > 1$. Both cases can be written as

$$\frac{w_\beta}{c_\alpha}(A\beta/2)^\theta.$$

This expression is strictly increasing in A . Solving for a ratio of one gives

$$A_c = \frac{2}{\beta} \left(\frac{c_\alpha}{w_\beta} \right)^{1/\theta}.$$

The comparisons on either side follow from the strict inequalities in the limit. If $A_c < 1$, every admissible $A \geq 1$ already lies above the cutoff. Equality gives a limiting ratio of one. $\square$

Proof of Proposition 3. At a fixed research state, D is unchanged. If the communicated sets satisfy $P \subseteq P'$, then

$$U(D, P') - U(D, P) = \mu(|P'| - |P|) = \mu|P' \setminus P|.$$

Every additional communicated concept therefore has marginal value μ . Applying the same inclusion to the AI record at each date weakly raises AI's value at that date and its ratio to the fixed human benchmark. Any strict improvement or change in a limiting threshold depends on the scale of the additional communication. $\square$

References

- S. Bubeck and E. Gwynne. Asymptotic behavior of the Eden model with positively homogeneous edge weights. arXiv:1508.05140v3, 2016.
- C. Carnehl and J. Schneider. A quest for knowledge. *Econometrica*, 93(2):623–659, 2025.
- A. Fogli and L. Veldkamp. Germs, social networks, and growth. *Review of Economic Studies*, 88(3):1074–1100, 2021.
- J. S. Gans. A quest for AI knowledge. Working Paper 33566, National Bureau of Economic Research, 2025.
- T. Gowers. Why I didn't sign the Fields medallists' letter. Gowers's Weblog, 2026. URL <https://gowers.wordpress.com/2026/09/17/why-i-didnt-sign-the-fields-medallists-letter/>. 17 September.
- Math and AI. A severe misalignment of AI in mathematics. Joint declaration, 2026. URL <https://mathandai.org/>. 11 September.
- OpenAI. Sharing AI progress in mathematics. Official announcement, 2026. URL <https://openai.com/index/sharing-ai-progress-in-mathematics/>. 6 October. Updated 15 October.
- G. Silverberg and B. Verspagen. A percolation model of innovation in complex technology spaces. *Journal of Economic Dynamics and Control*, 29(1–2):225–244, 2005.
- T. Tao. Mathematics in the age of AI. arXiv:2608.16753, 2026a. Essay based on the ICM public lecture.
- T. Tao. Math 2.0. Public lecture, Caltech, 2026b. URL <https://terrytao.wordpress.com/2026/10/10/math-2-0/>. 9 October. Slides posted 10 October.
- W. P. Thurston. On proof and progress in mathematics. *Bulletin of the American Mathematical Society*, 30(2):161–177, 1994.