A Quest for AI Knowledge

Joshua S. Gans*

March 5, 2025

Abstract

This paper examines how the introduction of artificial intelligence (AI), particularly generative and large language models capable of interpolating precisely between known data points, reshapes scientists' incentives for pursuing novel versus incremental research. Extending the theoretical framework of Carnehl and Schneider (2025), we analyse how decision-makers leverage AI to improve precision within well-defined knowledge domains. We identify conditions under which the availability of AI tools encourages scientists to choose more socially valuable, highly novel research projects, contrasting sharply with traditional patterns of incremental knowledge growth. Our model demonstrates a critical complementarity: scientists strategically align their research novelty choices to maximise the domain where AI can reliably inform decision-making. This dynamic fundamentally transforms the evolution of scientific knowledge, leading either to systematic "stepping stone" expansions or endogenous research cycles of strategic knowledge deepening. We discuss the broader implications for science policy, highlighting how sufficiently capable AI tools could mitigate traditional inefficiencies in scientific innovation, aligning private research incentives closely with the social optimum. *Journal of Economic Literature* Classification Numbers: O31, O33, D83.

Keywords. artificial intelligence, innovation, scientific research, knowledge accumulation, research incentives

*Rotman School of Management, University of Toronto and NBER. Thanks to Steven Callandar and Johannes Schneider for helpful discussions and to chatgpt-o3-mini-high and Claude 3.7 (at least when it was willing to work) for excellent research assistance. And, yes, I am aware there is something 'meta' about all this. Responsibility for all errors remains my own. Mailto:joshua.gans@utoronto.ca

1 Introduction

The rapid development of artificial intelligence (AI) is fundamentally reshaping the landscape of scientific discovery and innovation. A key feature of current AI systems, particularly generative models and large language models (LLMs), is their ability to synthesise and recombine existing knowledge to provide precise answers to queries posed by decision-makers (DMs). Many contemporary AI experts portray generative AI—especially reasoning-oriented large language models—as powerful knowledge synthesis tools that augment decision-making by recombining existing information rather than generating fundamentally novel ideas. These models are trained on vast corpora and effectively distill patterns and facts from their data, functioning much like massive, implicit knowledge bases. For instance, empirical studies have demonstrated that LLMs can recall and integrate factual knowledge to answer queries, at times even outperforming human specialists in predicting outcomes by identifying latent patterns in the literature (Ludwig and Mullainathan (2024); Toner-Rodgers (2024)).

Notably, these AI tools do not generate fundamentally new knowledge outside existing domains; rather, they excel at interpolation, precisely answering queries located within a defined operational range around known data points.[1] For instance, if the effects of a drug are known at doses of 10mg and 50mg, an AI tool can reliably predict outcomes at intermediate dosages (e.g., 30mg) but cannot reliably predict outcomes beyond the existing range (e.g., 100mg). Real-world examples include DeepMind's AlphaFold, which revolutionised protein structure prediction by accurately interpolating between known protein structures, and generative AI models that systematically explore combinatorial chemical spaces to propose novel compounds (DeepMind, 2020; Toner-Rodgers, 2024). Thus, AI's primary strength lies in enhancing decision-making precision through efficient recombination of existing information, rather than extending the domain of knowledge itself.

This paper explores how the availability of such AI tools impacts the incentives and

[1] Theoretically, scholars caution that today's generative models lack genuine originality or understanding–Bender et al. (2021) famously describe them as "stochastic parrots" that merely reassemble linguistic fragments without true comprehension. In a similar vein, Marcus (2020) contends that LLM outputs are essentially "pastiche machines," rephrasing known content while contributing little conceptually new material. Although some critics argue that the recombinative capacity of these models can occasionally yield seemingly novel connections Doshi and Hauser (2024), the prevailing view remains that their creative output emerges primarily from the interpolation of extant data. This perspective motivates our modelling assumption that AI operates predominantly within established knowledge domains to support decision-making.

strategic choices of scientists engaged in research. Decision-makers, who previously relied solely on manual searches and faced inherent uncertainty, can now use AI to rapidly access precise, reliable answers within specific knowledge domains. Our central research question is: *How does the availability of AI, capable of providing perfect decisions within defined domains, alter scientists' incentives when choosing between novel and incremental research projects?*

Our analysis builds upon the theoretical framework of Carnehl and Schneider (2025), who model scientific knowledge as a collection of discovered question-answer pairs and conceptualise scientific research as a costly search process, with scientists rewarded based on how their discoveries improve decision-making outcomes.[2] Traditionally, scientists balance novelty—pursuing distant, uncertain knowledge—with reliability, seeking incremental discoveries close to established knowledge. Without AI, this balancing act typically results in systematic underinvestment in highly novel but socially valuable research, producing incremental "ladder-like" patterns of knowledge growth.

Our paper contributes to this literature by explicitly modelling AI as a tool employed by decision-makers, significantly extending the Carnehl and Schneider (2025) framework. Unlike their analysis, which highlights systematic underinvestment in novel research, we identify specific conditions under which AI reshapes scientists' incentives towards pursuing more novel, socially valuable research. We find that AI substantially reshapes scientists' research incentives by introducing a strong complementarity between scientific novelty and decision-making effectiveness. Specifically, when scientists anticipate that their discoveries, combined with AI, can significantly enhance the precision and reliability of decision-making, they strategically target domains where their knowledge will maximally empower AI use by decision-makers. The underlying intuition is straightforward: scientists choose domains and novelty levels that amplify the value of AI-driven decision-making improvements, aligning their interests more closely with socially optimal innovation.

This analysis illuminates active debate within the economic literature on the im-

[2]The Carnehl and Schneider (2025) framework extends Callander (2011), who model trial-and-error search in product innovation using a Brownian motion framework. In Callander's model, a product's outcome is determined by its "distance" from known designs—greater distance entails higher uncertainty but also potential for greater reward. This structure, where learning is localized and uncertainty increases with novelty, is adapted by Carnehl and Schneider to explain scientists' choices between novel exploration and incremental knowledge deepening.

pact of AI on scientific innovation. Early contributions by Cockburn et al. (2019) and Agrawal et al. (2019) emphasise AI's role as a general-purpose technology that reduces uncertainty and facilitates recombinant innovation by lowering the cost of prediction, enabling novel connections between disparate ideas. Aghion et al. (2019) expanded this perspective, suggesting AI might catalyse a new wave of knowledge accumulation analogous to historical industrial revolutions, highlighting AI's dual role in increasing the efficiency of knowledge production and reallocating research effort across varying levels of novelty. Empirical studies, such as those by Toner-Rodgers (2024), provide evidence that AI-assisted research yields higher-quality outcomes and innovation output by complementing human expertise—handling routine combinatorial tasks and thereby freeing scientists to pursue more ambitious projects.

However, the existing literature has yet to fully reconcile two competing views regarding the dynamic interplay between AI and scientific knowledge evolution. On one hand, generative AI systems, as discussed by Doshi and Hauser (2024) and Ludwig and Mullainathan (2024), excel at synthesizing existing knowledge to facilitate breakthrough discoveries by efficiently connecting disparate ideas.[3] On the other hand, critics such as Maynard (2024) caution that heavy reliance on AI might narrow research agendas, pushing scientists toward safer, more conventional projects that the AI can readily process. Unlike prior contributions, our paper explicitly focuses on AI as a tool employed by decision-makers, examining how this changes scientists' incentives rather than directly affecting their productivity.

The analysis here yields a particularly striking and perhaps counterintuitive result: when sufficiently capable, AI tools actually encourage scientists to pursue more novel rather than incremental research. While one might expect AI—which excels at interpolation between known points—to substitute for novel research, we find the opposite effect. Precisely because AI can perfectly handle interpolation within established knowledge domains, the marginal value of incremental research diminishes substantially. Scientists respond by strategically targeting discoveries at the boundary of AI's operational capabilities, creating what we call a "stepping stone" pattern of knowledge expansion.

This mechanism directly counters the incrementalism identified by Carnehl and

[3]In so doing, this paper is related to a literature on the direction of scientific innovation (e.g., Acemoglu (2011), Bryan and Lemus (2017) and Hopenhayn and Squintani (2021)) and how AI may alter that (Gans, 2024a) This includes broader concerns regarding how the availability of AI may alter the data provided for future AI training; see, for instance, Gans (2024b).

Schneider by redirecting scientific effort toward more ambitious research projects. When AI efficiently manages knowledge deepening within known domains, scientists rationally shift their attention to expanding knowledge frontiers. The resulting pattern aligns remarkably well with socially optimal innovation, as scientists push knowledge boundaries in discrete jumps that maximise the domain where AI can subsequently operate. Thus, rather than reinforcing incrementalism, capable AI tools fundamentally resolve the tension between novelty and reliability that has traditionally led scientists to underinvest in transformative research.

In sum, our theoretical analysis clarifies precisely how AI reshapes scientific incentives towards greater novelty, revealing conditions under which AI can mitigate traditional inefficiencies. These findings hold significant implications for science policy, particularly in promoting socially optimal patterns of scientific discovery and innovation..

The remainder of the paper is organised as follows. Section 2 presents our theoretical model, explicitly integrating AI capabilities into the Carnehl and Schneider (2025) framework. Section 3 analyses the baseline case without AI to establish a benchmark for comparison. Section 4 examines how AI transforms scientists' strategic choices and the resulting knowledge evolution patterns. Section 5 provides welfare implications, including how AI can align private and social incentives. Section 6 concludes by discussing broader implications for research strategy and science policy in an AI-enhanced scientific landscape.

2 Model Setup

The model developed here builds on the framework introduced by Carnehl and Schneider (2025), who provide a microfounded approach to understanding the generation of scientific knowledge, its instrumental value, and the strategic research decisions scientists make. Their approach offers a suitable foundation to examine artificial intelligence as a toolkit employed by decision-makers, capable of effectively leveraging existing knowledge and enhancing its practical applicability. By extending their setup, we explicitly model how variations in AI quality and capability influence the underlying processes of scientific knowledge accumulation. In what follows, we summarize the central components of the Carnehl and Schneider (2025) framework and clearly define the operational capabilities attributed to AI tools within our analysis.

To that end, the primary agents in the model are (1) a representative decision-maker (DM) who uses knowledge (and if available AI) to assist in choosing actions by *searching* for answers to specific questions; and (2) scientists (or specifically, a representative scientist) who *research* the answers to specific questions that may be asked by decision-makers.

2.1 Knowledge and Questions

The universe of questions is presented by the real line $\mathbb{R}$. Each question $x \in \mathbb{R}$ has a unique answer $y(x) \in \mathbb{R}$. The answers to questions are modeled as realisations of a standard Brownian motion of a random variable $Y(x)$. This captures the idea that similar questions tend to have similar answers, with uncertainty increasing with distance.[4]

Knowledge at time t is the finite collection of known question-answer pairs, denoted by $\mathcal{F}_t = \{(x_i, y(x_i))\}_{i=1}^{k_t}$ with $x_1 < x_2 < \cdots < x_k$. This ordered set captures the results of previous research, and serves as the basis for forming posterior beliefs about answers to new questions.

There are several reasons for adopting this specification. First, modelling knowledge as a function indexed by $x \in \mathbb{R}$ enables us to treat questions as points on a continuum. This abstraction captures the idea that knowledge is built gradually and incrementally across a wide array of topics or dimensions. Second, the choice of a standard Brownian motion reflects a stylised notion of research where new findings, though uncertain, evolve in a continuous and unpredictable manner. Moreover, the Gaussian nature of Brownian motion simplifies the analysis since the conditional distributions (posterior beliefs) are also Gaussian.

Given existing knowledge $\mathcal{F}_t$, agents form conjectures about the answers to unknown questions. For any question x, the conjecture is a normal distribution with mean $\mu_x(\mathcal{F}_t)$ and variance $\sigma_x^2(\mathcal{F}_t)$, derived from the properties of Brownian motion:

$$Y(x \mid \mathcal{F}_t) \sim N\Big(\mu_x(Y \mid \mathcal{F}_t),\, \sigma_x^2(Y \mid \mathcal{F}_t)\Big). \tag{1}$$

Areas between known questions represent domains where conjectures have varying

[4]Formally, a standard Brownian motion $\{Y(x) : x \geq 0\}$ is a continuous-time stochastic process with the following properties: (1) $Y(0) = 0$ almost surely; (2) the process has independent increments; (3) for any $0 \leq s < t$, the increment $Y(t) - Y(s)$ is normally distributed with mean 0 and variance $t - s$; and (4) the sample paths of the process are continuous with probability one.

levels of precision. If x_i and x_{i+1} are adjacent known questions, the variance of the conjecture for a question $x \in [x_i, x_{i+1}]$ is given by:

$$\sigma^2(d; X) = \frac{d(X-d)}{X} \tag{2}$$

where $\sigma^2(d; X) = \sigma_x^2(Y \mid \mathcal{F}_k)$ and $d = x - x_i$ is the distance from the nearest known question and $X = x_{i+1} - x_i$ is the length of the area.

2.2 Decision-Makers and Value of Knowledge

A decision-maker (DM) uses available knowledge to address questions across the entire domain. For each question x, the DM must decide whether to take an action, $a(x)$, or abstain (denoted as $a(x) = \varnothing$). To make this decision, DM searches the corpus of existing knowledge and uses that knowledge to form a posterior belief as outlined in the previous subsection.

The payoff from taking an action is defined by the quadratic loss function:

$$u(a(x), x) = \begin{cases} 0, & \text{if } a(x) = \varnothing, \\ 1 - \dfrac{(a(x) - y(x))^2}{q}, & \text{if } a(x) \in \mathbb{R}. \end{cases}$$

where $q > 0$ is an exogenous parameter representing the importance of precision. In the Carnehl and Schneider (2025) framework, q captures the strength of the DM's outside option—that is, it reflects the relative attractiveness of forgoing the application of imprecise knowledge. A higher q implies that the alternative is more appealing, thereby requiring a greater degree of precision for the applied knowledge to yield a payoff above this fallback level.

Because of the quadratic loss structure, the optimal decision is to choose the posterior mean:

$$a^*(x \mid \mathcal{F}_t) = \mu_x(Y \mid \mathcal{F}_t)$$

provided that the uncertainty (quantified by the posterior variance) is sufficiently low. Formally, decision-makers act only when

$$\sigma_x^2(Y \mid \mathcal{F}_t) \leq q.$$

When this condition is met, the expected payoff from applying knowledge is:

$$1 - \frac{\sigma_x^2(Y|\mathcal{F}_t)}{q} \tag{3}$$

Several modelling choices are embedded in this decision rule. The use of quadratic loss is standard in decision theory. It not only simplifies analysis (since the posterior mean minimises expected quadratic loss) but also provides a smooth measure of how deviations from the true value affect payoffs. The condition $\sigma_x^2(Y \mid \mathcal{F}_t) \leq q$ ensures that decision-makers act only when they are sufficiently confident in the answer, capturing the realistic behaviour that a decision-maker will withhold action if uncertainty is too high, thereby avoiding costly mistakes.

The overall value of knowledge is:

$$v(\mathcal{F}_t) := \int_{-\infty}^{\infty} \max\left(1 - \frac{\sigma_x^2(Y|\mathcal{F}_t)}{q}, 0\right) dx \tag{4}$$

The benefits of a discovery at position x are the marginal increase in the value of knowledge:

$$V(x; \mathcal{F}_t) := v(\mathcal{F}_t \cup \{(x, y(x))\}) - v(\mathcal{F}_t) \tag{5}$$

A key insight of the model, as explored by Carnehl and Schneider (2025), is that the spatial (or conceptual) distribution of knowledge plays a critical role in determining the precision with which decision-makers can act. In particular, when the gap between adjacent knowledge points is X, the maximum posterior variance within that interval is $\frac{X}{4}$, which occurs at the midpoint. Consequently, decision-makers will act for every question in that interval if and only if $X \leq 4q$.

2.3 Scientist Choices

Scientists choose research projects in order to maximise the contribution of new knowledge toward improved decision-making. Thus, the assumption made by Carnehl and Schneider (2025) that the rewards to scientists are proportional to the direct impact of their research is maintained; that is, $V(x; \mathcal{F}_t)$ as derived above.

In each period, a scientist selects a question x and a search intensity that determines the probability $\rho \in [0, 1]$ of discovering the answer. Denote by X the length of the research area (or gap) within which the new question x lies, and let $d \in [0, X/2]$

denote the distance of x from the nearest known point so that d serves as a measure of *novelty*. Novelty is valuable in that a discovery farther from existing knowledge is more likely to fill a significant gap and thereby reduce uncertainty in a previously underexplored area. However, higher novelty comes at a cost.

Specifically, recall that the posterior variance at a distance d in an interval of length X is given by $\sigma^2(d;X) = \frac{d\,(X-d)}{X}$, which increases (up to a point) with d. A higher variance not only reduces the immediate benefit (since decision-makers act only when the uncertainty is low) but also makes the research process riskier. This is captured in the scientist's cost function which is assumed to take the form:

$$c(\rho, d; X) = \eta\tilde{c}(\rho)\sigma^2(d;X) \tag{6}$$

where $\eta > 0$ is a cost parameter, $\tilde{c}(\rho) = (\text{erf}^{-1}(\rho))^2$ is an increasing, convex function of the success probability, and $\sigma^2(d;X)$ is the variance of the conjecture about the answer.[5]

The scientist's payoff is: $u_R(d, \rho; X) = \rho V(d;X) - \eta\tilde{c}(\rho)\sigma^2(d;X)$. Thus, the scientist's problem is to choose X, d, and ρ to maximise

$$\max_{X\in\{X_0,\ldots,X_k\}} \; \max_{d\in[0,X/2],\,\rho\in[0,1]} \quad \rho\, V(d;X) - \eta\, c(\rho, d; X),$$

where $\eta > 0$ is a scaling factor that converts research cost into utility units.

The optimal choices of the scientist involve a multi-dimensional trade-off. A larger d implies that the research targets a frontier area—one that is more distant from existing knowledge. Such a discovery can yield a higher $V(d;X)$ because it fills an important gap and substantially reduces uncertainty in an otherwise sparse region. However, as d increases, the posterior variance $\sigma^2(d;X)$ also increases, thereby reducing the probability that the decision-maker would otherwise act and increasing the

[5] Carnehl and Schneider (2025) provide a microfoundation for this cost function based on a model of optimal search. They conceptualise research as a process where the scientist chooses a sampling interval $[a, b]$ in which to search for the answer. The scientist discovers the answer if and only if the true answer $y(x)$ falls within this interval. With a Gaussian posterior distribution for the answer, the optimal search strategy is to center the interval around the mean of the posterior. The probability of success ρ corresponds to the probability mass of the posterior distribution covered by the sampling interval, which is related to the inverse error function. The quadratic form $(\text{erf}^{-1}(\rho))^2$ emerges because achieving a higher success probability requires widening the sampling interval, with the cost increasing quadratically with the interval width. The term $\sigma^2(d;X)$ captures the idea that search is more costly when the variance is higher, as the sampling interval must be wider to achieve the same probability of success.

cost $c(\rho, d; X)$.

The variable ρ reflects the probability that the research successfully yields a new discovery. A higher ρ directly increases the expected benefit $\rho\, V(d; X)$ but enters the cost function nonlinearly through the term $\text{erf}^{-1}(\rho)$, making high success probability costly. Thus, scientists must balance the desire for a high probability of success against the escalating cost of achieving it.

Finally, the parameter X captures the width of the region under investigation. A larger X is indicative of a sparser region of knowledge, where there is more scope for a novel discovery to reduce uncertainty. However, if X is too large, then even moderate values of d could yield prohibitively high posterior variances, leading to increased research costs. Conversely, a very small X implies that the knowledge base is already dense, and the marginal benefit from an additional discovery is low.

2.4 Capabilities of the AI Tool

Before introducing our AI tool, it is important to understand a key distinction from Carnehl and Schneider (2025) regarding two fundamental modes of knowledge accumulation: *deepening* versus *expanding* knowledge.

- **Deepening knowledge** refers to the process of discovering answers to questions that lie between two existing knowledge points $(x_i, y(x_i))$ and $(x_{i+1}, y(x_{i+1}))$. When a scientist deepens knowledge, they fill in gaps within already established domains, increasing the density of knowledge in that region. For example, if we know the effects of a drug at doses of 10mg and 50mg, deepening knowledge would involve determining its effect at 30mg.

- **Expanding knowledge**, in contrast, involves discovering answers to questions that lie beyond the current frontiers of knowledge. This pushes the boundaries of knowledge outward into previously unexplored territories. Using the same example, expanding knowledge might involve testing the drug at 100mg, beyond the highest known dose.

In Carnehl and Schneider (2025), scientists may choose either strategy depending on the relative benefits and costs. The benefits of deepening are highest when the gap between existing knowledge points is large, as this reduces uncertainty in regions where predictions would otherwise be poor. The benefits of expanding knowledge

are generally constant but face increasing costs as one moves farther from existing knowledge.

We now extend this framework by introducing an AI tool that can be used by DMs with the following capability:

Definition (AI Capability)**.** *The AI tool provides a perfectly reliable decision (yielding a payoff of 1) for any question x located between two known points $(x_i, y(x_i))$ and $(x_j, y(x_j))$, if and only if the distance between these points satisfies $|x_j - x_i| \leq X_A$, where $X_A > 0$ is an exogenously given parameter representing the AI's operational range. If $|x_j - x_i| > X_A$, the AI provides no reliable decisions for questions in this range.*

This definition formalises a fundamental characteristic of modern AI systems: they excel at interpolation within dense regions of knowledge but struggle with extrapolation beyond established frontiers. Here we capture this notion by assuming that the AI tool can only *deepen* knowledge, not *expand* it. (See Figure 1).

When the distance between two adjacent knowledge points is less than or equal to X_A, the AI can provide perfect decisions for all questions in the interval between them. This means that for any question $x \in [x_i, x_j]$ where $|x_j - x_i| \leq X_A$, the decision-maker can use the AI to generate a choice of $a(x)$ that is perfect, yielding the maximum possible payoff of 1.

Critically, this capability creates a binary outcome: either the AI can perfectly answer questions between two points (if they're close enough), or it cannot provide any improvement over standard conjectures (if they're too far apart). This assumption is made for simplicity and generates a sharp distinction associated with using the AI tool. The AI tool can generate a perfectly reliable decision for questions within the domain, X_A, and not otherwise. This allows us to more clearly see the impact of AI that could be 'smoothed' over if AI generated more reliable but not perfectly reliable decisions and could generate decisions with some reliability just outside of the domain, X_A. For instance, we could imagine that the AI tool reduced the variance of the error term faced by DM within the domain X_A and reducing it by less outside that domain. Adjusting the model in this way would not fundamentally change the qualitative results derived below.

A real-world example helps illustrate this concept. Consider the problem of protein folding, where scientists aim to predict a protein's three-dimensional structure from its

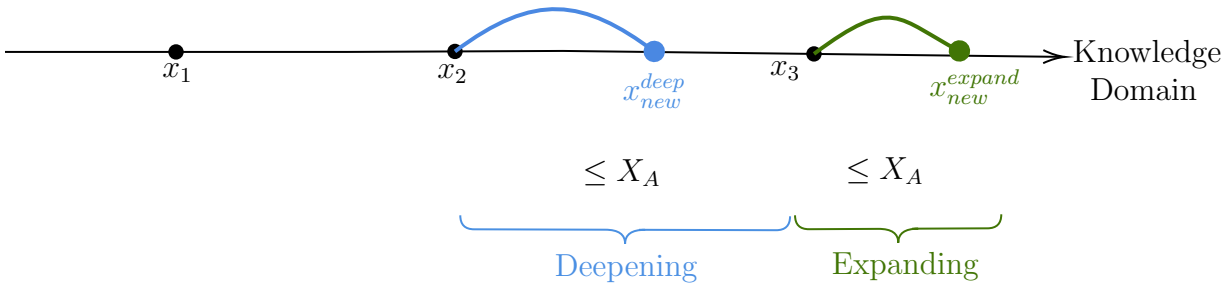


Figure 1: **Illustration of deepening versus expanding knowledge**. Blue indicates deepening knowledge by discovering x_{new}^{deep} between existing points x_2 and x_3. Green indicates expanding knowledge by discovering x_{new}^{expand} beyond the frontier point x_3. The brackets depict the AI's operational range X_A from each knowledge point.

amino acid sequence. Traditionally, scientists would need to experimentally determine structures for many similar proteins to make reliable predictions about new ones. With AI tools like AlphaFold, if we know the structures of two somewhat related proteins (separated by a "distance" less than X_A in the space of amino acid sequences), the AI can accurately predict the structure of any protein between them without requiring additional experimental data. However, for proteins that are too distant from known structures (beyond the frontier), the AI's predictions become unreliable, and new experiments are needed.

3 Analysis Without AI

Having established our theoretical framework—which models knowledge as a collection of question-answer pairs, scientific research as a costly search process, and AI tools as providing perfect answers within domains of specific length—we now analyse how scientists make optimal research decisions in the absence of AI. This baseline analysis serves two critical purposes. First, it provides a benchmark against which we can evaluate the impact of AI tools on research incentives. Second, understanding the inherent trade-offs scientists face between novelty and reliability helps illuminate how AI fundamentally alters these trade-offs. By first establishing how knowledge evolves without AI intervention, we can precisely identify the mechanisms through which AI transforms scientific discovery and the resulting knowledge landscape.

3.1 The Benefits of Discovery

The benefits of a discovery depend crucially on its location relative to existing knowledge. Carnehl and Schneider (2025) derive a comprehensive benefit function that captures how a new discovery improves decision-making. Here, we provide an intuitive explanation of this function.

Abusing notation, for a discovery at distance $d \in [0, X/2]$ from the nearest known point in an area of length $X \in \mathbb{R} \cup \{\infty\}$, the benefit function is:

$$V(d;X) := \frac{1}{6q}\Big(2X\sigma^2(d;X) + \mathbf{1}_{d>4q}\sqrt{d}(d-4q)^{3/2} + \mathbf{1}_{X-d>4q}\sqrt{X-d}(X-d-4q)^{3/2} - \mathbf{1}_{X>4q}\sqrt{X}(X-4q)^{3/2}\Big),$$

where $V(d;\infty) := \lim_{X\to\infty} V(d;X)$. While this expression appears complex, it has a clear interpretation. The first term captures the direct reduction in uncertainty from adding a new knowledge point. The indicator terms reflect scenarios where the DM would not act before the discovery (due to high uncertainty) but would act afterwards. Specifically, the term with $1_{d>4q}$ represents regions to the left of the new discovery that become actionable; the term with $1_{X-d>4q}$ covers regions to the right of the new discovery that become actionable; and the term with $1_{X>4q}$ subtracts regions that were already actionable before the discovery.

As mentioned earlier, we can distinguish between scientists deepening knowledge versus expanding knowledge. When deepening knowledge—that is, when a discovery is made within an existing interval of length X—the net benefit from replacing the interval with two shorter ones is given by:

$$V(d;X) = \frac{d(X-d)}{3q}. \tag{7}$$

This expression is maximised at $d = \frac{X}{2}$; in particular, when $X = 6q$, the optimal deepening choice occurs at $d = 3q$, splitting the interval into two equal parts of length $3q$ and yielding the highest benefit. When the scientist focuses only on expanding knowledge (when there is no upper bound on X), this simplifies to:

$$V(d;\infty) = d - \frac{d^2}{6q} \tag{8}$$

for $d \leq 3q$, which increases until $d = 3q$ and then decreases. As these distinct benefits show, when expanding knowledge, scientists increase d to push into unexplored territory, with benefits that initially increase (as uncertainty reduction covers more area) but eventually decrease (as distance makes research more costly). When deepening knowledge, scientists choose a position within an existing gap, balancing the benefits of creating two smaller areas with more precise conjectures against the costs.

3.2 Optimal Scientist Choices

To understand scientist behaviour, it is useful to analyse separately the cases of expanding knowledge (pushing into unexplored territory) versus deepening knowledge (filling gaps between known points). This separation clarifies the fundamental trade-offs scientists face when choosing research questions.

3.2.1 Optimal Distance for Expanding Knowledge

When expanding knowledge, scientists select a distance $d^{\eta}(\infty)$ from the current frontier that optimally balances novelty against precision. The notation $d^{\eta}(\infty)$ indicates that this is the optimal distance given cost parameter η in an unbounded research area (denoted by ∞). Using the earlier benefit function for expanding knowledge (8), the first-order condition with respect to d yields:

$$\rho \frac{\partial V(d;\infty)}{\partial d} = \eta \tilde{c}(\rho) \frac{\partial \sigma^2(d;\infty)}{\partial d} \tag{9}$$

Since $\frac{\partial V(d;\infty)}{\partial d} = 1 - \frac{d}{3q}$ and $\frac{\partial \sigma^2(d;\infty)}{\partial d} \approx 1$ for expanding knowledge, this simplifies to:

$$\rho \left(1 - \frac{d}{3q}\right) = \eta \tilde{c}(\rho) \tag{10}$$

Analysis of this condition reveals that the optimal distance $d^{\eta}(\infty)$ lies in the interval $(2q, 3q]$. The upper bound of $3q$ is clear since the marginal benefit $\frac{\partial V(d;\infty)}{\partial d}$ becomes zero at $d = 3q$. The lower bound of $2q$ emerges from solving the system of first-order conditions for both d and ρ simultaneously; for $d < 2q$, the scientist would need to select an unrealistically high success probability to satisfy both optimality conditions,

resulting in a joint solution that pushes d above $2q$ even with positive research costs.[6]

Intuitively, scientists face a critical trade-off when expanding knowledge: venturing farther into unexplored territory (increasing d) initially increases benefits by reducing uncertainty over wider regions, but beyond a certain point (specifically, $3q$), the marginal benefits decline while costs continue to rise. When $d < 2q$, the research is too incremental—the discovery would provide minimal new insights despite being relatively easy to achieve. When $d > 3q$, the research becomes excessively speculative—the increased difficulty and uncertainty outweigh the potential gains from greater novelty.

This optimal range $(2q, 3q]$ embodies what might be called the "Goldilocks principle" of scientific research: discoveries should be novel enough to significantly advance knowledge, but not so distant that they become prohibitively difficult to achieve reliably.

3.2.2 Optimal Distance for Deepening Knowledge

When deepening knowledge within an area of length X, the benefit function changes earlier stated as (7) applies. The first-order condition with respect to d gives:

$$\frac{\partial V(d;X)}{\partial d} = \frac{X - 2d}{3q} = 0 \tag{11}$$

This yields $d = X/2$ as the optimal choice for intermediate-length areas.

The trade-offs for deepening knowledge differ significantly from those for expanding knowledge. When selecting a position within an existing gap, scientists balance the quality of the two new regions created by their discovery. Choosing the midpoint ($d = X/2$) creates two equal regions, optimally distributing the reduction in uncertainty. This maximises the overall improvement in decision-making accuracy across the entire gap.

However, for very long areas (when $X > 8q$), a different trade-off emerges. Rather

[6] To see why $d > 2q$, we solve the joint first-order conditions. The FOC for ρ gives $d - \frac{d^2}{6q} = \eta\tilde{c}'(\rho)d$, which simplifies to $1 - \frac{d}{6q} = \eta\tilde{c}'(\rho)$. Substituting into the FOC for d, we get $\rho(1 - \frac{d}{3q}) = \eta\tilde{c}(\rho)$. When we combine these equations, we obtain $\frac{\tilde{c}(\rho)}{\tilde{c}'(\rho)} = \rho \cdot \frac{1-\frac{d}{3q}}{1-\frac{d}{6q}}$. Given the properties of $\tilde{c}(\rho) = (\text{erf}^{-1}(\rho))^2$, the left-hand side approaches zero as ρ approaches 1. For the equation to hold with realistic values of ρ, the fraction on the right must be sufficiently large, which requires $d > 2q$. More precisely, as η increases, the optimal ρ decreases, which necessitates $d > 2q$ to maintain the optimality conditions.

than evenly dividing a large gap, scientists prefer positions that ensure at least one region has sufficiently low variance to enable reliable decision-making. This represents a shift from "even coverage" to "targeted precision" as gaps grow larger—creating one highly usable region becomes more valuable than modestly improving two regions that remain too uncertain for practical use.

3.2.3 Success Probability

Regardless of whether expanding or deepening knowledge, the scientist's choice of success probability ρ is determined by the selected distance d. The first-order condition with respect to ρ is:

$$V(d;\infty) = \eta\tilde{c}'(\rho)d \tag{12}$$

for expanding knowledge, or the analogous expression for deepening knowledge.

The scientist faces a fundamental trade-off when selecting ρ: increasing the probability of successful discovery directly enhances expected benefits, but achieving higher reliability requires increasingly costly investments. Given that $\tilde{c}(\rho) = (\text{erf}^{-1}(\rho))^2$ is convex and increases without bound as ρ approaches 1, scientists optimally select $\rho^\eta \equiv \rho^*(\eta) < 1$.

Importantly, the optimal success probability depends directly on the chosen distance d. More distant or novel research (larger d) increases both the potential benefits of discovery and the difficulty of achieving reliable results. This creates a natural correlation: more ambitious research projects (those with larger d) typically operate with lower success probabilities, while more incremental projects can justify higher investments in reliability.

3.3 The Evolution of Knowledge

The following proposition characterises the conditions under which scientists direct research towards expanding rather than deepening knowledge.

Proposition 1 (Deepening vs. Expanding)**.** *There exists a threshold $X_c^\eta < X_c^0 \approx 4.338q$ such that scientists expand knowledge if and only if all bounded areas have length less than X_c^η.*

The proofs for all results are in the appendix. This result reveals the strategic calculation scientists make when deciding whether to deepen existing knowledge or expand

into new territories. When all knowledge gaps are relatively small ($X < X_c^\eta$), the marginal benefit of exploring entirely new questions exceeds that of refining answers within existing domains. However, once any gap becomes sufficiently wide, the value of reducing uncertainty within that sparse region outweighs the benefits of frontier exploration.

This threshold behaviour creates a natural rhythm in the evolution of knowledge: periods of expansion are followed by periods of consolidation once gaps become too large; something Carnehl and Schneider (2025) refer to as a "research cycle." The threshold X_c^η decreases with research costs, indicating that in environments where research is more expensive, scientists more readily shift toward deepening existing knowledge rather than pursuing new frontiers.

An important finding from Carnehl and Schneider (2025) is that knowledge tends to evolve in a "ladder structure":

Proposition 2 (Ladder Structure)**.** *Starting from a single knowledge point, scientists always choose to expand knowledge, leading to a step-by-step expansion of the knowledge frontier.*

This occurs because when a scientist expands knowledge, the newly created area has length equal to the expansion distance $d^\eta(\infty) < X_c^\eta$. Since this area is shorter than the threshold for deepening, the next scientist will also choose to expand, continuing the pattern.

From a social welfare perspective, this ladder structure may be suboptimal. Carnehl and Schneider (2025) demonstrate that scientists tend to choose distances that are too short and success probabilities that are too low compared to the social optimum. This occurs because scientists do not internalise the positive externalities their discoveries create for future research. For instance, the social planner might prefer "moonshots" - more distant discoveries that, while suboptimal in the short run, create better research opportunities for future generations. Such moonshots can break the ladder structure by creating knowledge gaps large enough to incentivise deepening, leading to research cycles that ultimately generate more valuable knowledge.

Understanding these baseline dynamics is crucial for appreciating how AI tools transform research incentives, as we will explore in the following sections. The key tensions between deepening and expanding knowledge, the critical distance thresholds, and the social welfare implications provide the foundation for analyzing AI's impact on scientific progress.

4 Outcomes with AI

The analysis thus far reveals that without AI, scientists tend to make research choices that result in a ladder-like pattern of knowledge accumulation with systematic underinvestment in novelty. This incremental approach reflects scientists' rational response to the fundamental trade-off between pushing knowledge boundaries and maintaining research reliability. We now turn to the paper's core contribution: understanding how the introduction of AI capabilities transforms these dynamics. By explicitly modelling AI as a tool that provides perfect decisions within an operational range of length X_A, we can examine how this technological capability reshapes scientists' strategic choices, the resulting knowledge evolution patterns, and, ultimately, social welfare. As we will demonstrate, AI creates novel complementarities between human research and machine capabilities that fundamentally alter both the quantitative and qualitative aspects of scientific progress.

This section considers how the presence of an AI tool affects the scientist's problem and the evolution of knowledge. The key insight is that a new discovery not only creates direct value by revealing the answer to a question but also potentially enables the AI to provide perfect decisions when combined with existing knowledge points if they are within the AI's operational range. That is, when a scientist discovers the answer to a new question $x' \notin [x_i - X_A, x_i + X_A]$, this not only adds to knowledge directly but also potentially extends the domain where the AI can provide perfect decisions. If the new discovery creates a gap with an existing knowledge point that is less than or equal to X_A, then the AI suddenly becomes operational in this newly created interval. Critically, if the new point is too far from existing knowledge (beyond distance X_A), the AI cannot bridge this gap and provides no additional benefit beyond the standard improvement in conjectures.

As will be shown, this dynamic creates a novel complementarity between human research and AI capabilities: scientists can strategically target discoveries that maximise the domain where AI can subsequently operate, potentially leading to different research incentives than in the original Carnehl and Schneider (2025) framework. The binary nature of AI capability—either providing perfect decisions or no additional benefit—creates distinct regimes of optimal research behaviour that we will examine in detail.

4.1 Modified Benefit Function

With AI, the benefit function for a scientist is fundamentally altered. The AI's capability creates distinct scenarios for both expanding and deepening knowledge.

Lemma 1 (Modified Benefit Function)**.** *The benefit function when a scientist discovers a new answer depends on whether the scientist is expanding or deepening knowledge.* ***For expanding knowledge*** *(discovering a question at distance d beyond the current frontier point x_0):*

$$V_E(d; X_A) = \begin{cases} d & \text{if } d \leq X_A \\ V(d; \infty) = d - \frac{d^2}{6q} & \text{if } d > X_A \text{ and } d \leq 3q \end{cases} \tag{13}$$

where the function continues as in the no-AI case for $d > 3q$. ***For deepening knowledge*** *(discovering a question at position d within an interval $[x_0, x_1]$ of length $X = x_1 - x_0$):*

$$V_D(d; X, X_A) = \begin{cases} X & \text{if } d \leq X_A \text{ and } X - d \leq X_A \\ d & \text{if } d \leq X_A \text{ and } X - d > X_A \\ X - d & \text{if } d > X_A \text{ and } X - d \leq X_A \\ V(d; X) = \frac{d(X-d)}{3q} & \text{if } d > X_A \text{ and } X - d > X_A \end{cases} \tag{14}$$

Lemma 1 formalises precisely how the introduction of AI alters the scientist's benefit function by embedding the binary operational nature of AI capability. This modification is crucial as it directly sets the stage for subsequent results, particularly illustrating how scientists' incentives shift due to AI-induced complementarities. The lemma clarifies conditions under which discoveries significantly enhance AI effectiveness, thus motivating why strategic placement of discoveries becomes essential to maximising knowledge benefits.

For expanding knowledge, $V_E(.)$ captures the transformative effect of AI on research incentives. If the scientist discovers an answer at distance $d \leq X_A$ from the nearest existing knowledge point, the AI can provide perfect decisions over the entire new interval, generating a benefit of d. However, if the scientist ventures beyond the AI's capability ($d > X_A$), the AI cannot connect this new point with existing knowledge, and the benefit reverts to the original function from the no-AI case.

For deepening knowledge, $V_D(.)$ reveals a powerful "2-for-1" opportunity. Consider an interval $[x_0, x_1]$ with length $X > X_A$ (beyond the AI's operational range). By strategically placing a new discovery at position d such that both $d \leq X_A$ and $X - d \leq X_A$, the scientist enables the AI to operate over the entire original interval. Both resulting subintervals $[x_0, x_0 + d]$ and $[x_0 + d, x_1]$ fall within the AI's capability, generating a total benefit of X. This creates a particularly compelling incentive when $X > 2X_A$: without the strategic placement, the AI couldn't operate on any part of the interval, but with a single discovery near the midpoint, the AI can suddenly provide perfect decisions throughout the entire region. When only one of the resulting intervals falls within the AI's capability (as in the second and third cases), the benefit equals the length of that interval.

The AI's capability fundamentally alters scientists' incentives, particularly for deepening knowledge in areas where $X > X_A$. This creates a new complementarity between human research and AI capabilities that was absent in the no-AI framework.

4.2 Scientist Choices with Expanding Knowledge

We first analyse how the AI tool affects scientist choices when focusing exclusively on expanding knowledge. In so doing, the following definition is of use:

Definition 1 (AI Capability Threshold for Expanding Knowledge)**.** The threshold X_E^* represents the critical AI capability level at which scientists transition from ignoring the AI capability boundary to precisely targeting it. Formally:

$$X_E^* = \frac{U_2^*}{\rho_1^* - \eta\tilde{c}(\rho_1^*)} \tag{15}$$

where U_2^* is the expected utility from the no-AI expansion strategy, ρ_1^* is the success probability that satisfies $1 = \eta\tilde{c}'(\rho_1^*)$, and the denominator represents the net marginal benefit per unit distance when expanding exactly to the AI capability boundary.

Proposition 3 (Optimal Expanding Choices with AI)**.** *When scientists focus solely on expanding knowledge in the presence of an AI tool with capability X_A, there exist two possible regimes based on distance choice:*

- *Regime 1E: Choose $d \leq X_A$, obtaining benefit $V_E(d) = d$. The scientists choose $d_1^* = X_A$ (exactly the AI capability boundary) and ρ_1^* satisfying $1 = \eta\tilde{c}'(\rho_1^*)$.*

- *Regime 2E: Choose $d > X_A$, with benefit $V_E(d) = d - \frac{d^2}{6q}$ for $d \leq 3q$. The scientists choose $d_2^* = d^\eta(\infty)$ (the same as in the no-AI case) and ρ_2^* satisfying $1 - \frac{d^\eta(\infty)}{6q} = \eta\tilde{c}'(\rho_2^*)$.*

There exists a threshold X_E^* such that: for $X_A > (<) X_E^*$, scientists choose Regime 1E (2E). For $X_A > X_E^*$, the optimal success probability ρ_1^* exceeds the no-AI optimal $\rho^\eta(\infty)$.

The intuition behind this result is straightforward: scientists face a strategic decision when expanding knowledge. When the AI's capability is limited (below threshold X_E^*), scientists ignore the AI constraint and choose the same distance as they would without AI, accepting that the AI cannot operate in the newly created region. This occurs because the benefits of pursuing a more novel discovery outweigh the benefits of enabling AI operation.

However, when the AI's capability exceeds the threshold X_E^*, scientists dramatically shift their strategy, choosing to expand knowledge by exactly the distance the AI can bridge ($d = X_A$). This creates a perfect complementarity between human research and AI capabilities: humans systematically expand the frontier of knowledge in discrete jumps that precisely match the AI's operational range, maximising the domains where perfect decisions become possible.

Moreover, in this regime, scientists increase their success probability compared to the no-AI case. This occurs because the marginal benefit of success increases when discoveries enable perfect AI decisions, justifying higher investments in reliability despite the increased costs.

As an example, consider a scientist engaged in drug discovery exploring drug dosages. Without AI, they might optimally test a dose 2.5 times the current maximum (corresponding to $d^\eta(\infty) \approx 2.5q$). However, with an advanced AI system that can accurately predict effects within a range of 3 times the current maximum ($X_A = 3q > X_E^*$), scientists would instead test exactly at that 3x boundary and invest more resources to ensure experimental success. This creates an expanding ladder of tested dosages, with each new discovery precisely at the boundary of AI's predictive capability.

4.3 Scientist Choices with Deepening Knowledge

When scientists can deepen knowledge by placing discoveries within existing gaps, the AI tool creates powerful new incentives.

Proposition 4 (Optimal Deepening Choices with AI)**.** *When scientists consider deepening knowledge within an interval* $[x_0, x_1]$ *of length* $X = x_1 - x_0$ *in the presence of an AI tool with capability* X_A*:*

(a) For intervals where $X \leq 2X_A$*, scientists choose* $d^* = \min(X/2, X_A)$ *if* $X \leq 2X_A$ *and* ρ_D^* *satisfying* $1 = \eta \tilde{c}'(\rho_D^*)$*.*

(b) For intervals where $X > 2X_A$*, if strategic AI-enabling placement is possible, position* d^* *that creates at least one subinterval within AI capability; otherwise, position* d^* *as in the no-AI case.*

(c) The expected benefit from optimal deepening with AI strictly exceeds the no-AI benefit when $X_A > 0$ *and* $X \in (X_A, 2X_A]$

The key insight is that the AI tool creates a powerful incentive for strategic placement of discoveries when deepening knowledge. For intervals with length $X \in (X_A, 2X_A]$, scientists can transform a region where the AI initially cannot operate (since $X > X_A$) into two regions where the AI can operate perfectly (since both $d \leq X_A$ and $X - d \leq X_A$). This bridge-building strategy generates extraordinary value compared to the no-AI case.

For example, if $X = 1.8X_A$, the optimal position is $d = 0.9X_A$, creating two equal subintervals within the AI's capability. This generates a benefit of $X = 1.8X_A$, substantially exceeding the no-AI benefit of approximately $\frac{X^2}{12q}$. But even for larger intervals ($X > 2X_A$), scientists may strategically position discoveries to create at least one subinterval where the AI can operate, rather than following the no-AI optimal strategy of placing at the midpoint.

Continuing our drug discovery example, imagine scientists have tested dosages of 10mg and 50mg, creating a gap of 40mg that exceeds the AI's capability to interpolate (say $X_A = 25$mg). Rather than testing at the midpoint (30mg) as they might without AI, they would strategically test at 25mg. This creates one interval (10mg to 25mg) where the AI can provide perfect predictions, maximising the range where physicians can confidently prescribe intermediate dosages.

4.4 Evolution of Knowledge with AI

The combination of altered incentives for both expanding and deepening knowledge fundamentally changes how scientific knowledge evolves over time. To understand this evolution, we first characterise the threshold that determines whether scientists prefer expanding or deepening knowledge in the presence of AI.

When scientists choose between deepening knowledge in an interval of length X versus expanding knowledge with AI, they compare the expected utilities:

$$u_R^D(X; X_A) = \max_{d \in [0, X/2], \rho \in [0,1]} \rho V_D(d; X, X_A) - \eta \tilde{c}(\rho) \sigma^2(d; X) \tag{16}$$

and

$$u_R^E(X_A) = \max_{d > 0, \rho \in [0,1]} \rho V_E(d; X_A) - \eta \tilde{c}(\rho) d \tag{17}$$

From our previous analysis, when AI capability exceeds the threshold X_E^*, the optimal expanding choice is $d = X_A$ with corresponding success probability ρ_1^* satisfying $1 = \eta \tilde{c}'(\rho_1^*)$. This yields expected utility:

$$u_R^E(X_A) = X_A[\rho_1^* - \eta \tilde{c}(\rho_1^*)] \tag{18}$$

For deepening knowledge, the expected utility depends on the relationship between X and X_A. When $X \in (X_A, 2X_A]$, scientists can strategically position discoveries to enable the AI across the entire interval, yielding substantially higher benefits than in the no-AI case. This makes deepening more attractive for a wider range of interval lengths.

The threshold $X_c^{\eta,A}$ is defined as the interval length at which these utilities are equal:

$$u_R^D(X_c^{\eta,A}; X_A) = u_R^E(X_A) \tag{19}$$

This threshold is strictly smaller than the corresponding threshold without AI ($X_c^{\eta,A} < X_c^{\eta}$) because the AI enhances the benefits of deepening knowledge by enabling perfect decisions in regions where strategic discoveries are placed. As AI capability X_A increases, the threshold $X_c^{\eta,A}$ decreases even further, making deepening preferable for smaller and smaller intervals.

With this threshold established, we can now precisely characterize how knowledge evolves with AI:

Proposition 5 (Ladder Structure with AI)**.** *When AI capabilities exceed threshold X_E^*, the evolution of knowledge from a single point follows a distinctive pattern:*

(a) If $X_A < X_c^{\eta,A}$, scientists create a uniform "stepping stone" ladder by expanding knowledge in discrete steps of exactly X_A, where $X_c^{\eta,A}$ is defined as:

$$X_c^{\eta,A} = \inf\{X > 0 : u_R^D(X; X_A) \geq u_R^E(X_A)\} \tag{20}$$

with $u_R^D(X; X_A)$ and $u_R^E(X_A)$ representing the expected utilities from deepening and expanding knowledge, respectively. This pattern continues indefinitely, with each scientist systematically extending the frontier by precisely the AI's operational range.

(b) If $X_A > X_c^{\eta,A}$, after the first expansion, scientists switch to deepening knowledge in the created interval, placing discoveries strategically to maximise AI effectiveness. This creates a cyclical pattern where periods of expansion are followed by periods of strategic deepening.

(c) As X_A increases, the threshold $X_c^{\eta,A}$ decreases, increasing the prevalence of deepening versus expanding.

This result reveals how AI fundamentally reshapes the evolution of knowledge. Without AI, knowledge typically evolves in a stepwise pattern with consistent distances determined by the cost-benefit trade-off faced by each scientist. With AI, when its capabilities exceed the threshold X_E^*, scientists systematically expand the frontier in discrete jumps of exactly X_A, creating a distinctive stepping-stone pattern that precisely matches the AI's operational range.

The key insight is that scientists optimally choose to expand knowledge by exactly the length of domain where AI can operate. When $X_A < X_c^{\eta,A}$, this stepping-stone pattern continues indefinitely, creating a highly regular ladder structure with uniform intervals of exactly X_A. This contrasts with the no-AI case, where the distance between adjacent knowledge points is determined by research costs and uncertainty parameters, not by a technological capability boundary.

When $X_A > X_c^{\eta,A}$, a more complex pattern emerges. After scientists expand to create an interval of length X_A, they find it more beneficial to deepen knowledge within this interval rather than continue expanding. By strategically placing discoveries, they create subintervals that all fall within the AI's capability, enabling perfect

decisions throughout the original interval. This creates a cyclical pattern where expansion is followed by strategic deepening, leading to an increasingly dense knowledge structure in certain regions.

Returning to our drug discovery example, imagine that AI systems can reliably predict drug effects for dosage intervals of at most 25mg (i.e., $X_A = 25\text{mg}$). When $X_A < X_c^{\eta,A}$, scientists would establish a perfectly regular pattern of tested dosages at exactly 25mg intervals: 0mg, 25mg, 50mg, 75mg, 100mg, and so on. Each scientist systematically extends the frontier by precisely 25mg, the maximum distance where AI can immediately apply the new knowledge.

However, if $X_A > X_c^{\eta,A}$ (perhaps because testing costs are high, making deepening more attractive), scientists would first test at 25mg, but then would strategically test an intermediate dosage like 12.5mg rather than extending to 50mg. This creates two intervals of 12.5mg each, both within the AI's capability, enabling perfect predictions throughout the entire 0mg to 25mg range. This strategic deepening maximises the regions where physicians can confidently prescribe intermediate dosages. Future scientists might then expand from 25mg to 50mg, followed by similar strategic deepening, creating a cyclical pattern of frontier extension followed by knowledge consolidation.

This methodical, AI-guided evolution of knowledge represents a fundamental shift from traditional research patterns, with scientists systematically pushing boundaries in ways that maximise the immediate applicability of AI tools to improve decision-making. In the (online) appendix, there is further analysis of the potential phase transition that might be observed in research as AI capabilities from $X_A < X_E^*$ to $X_A > X_E^*$ representing a jump in the domain researched.

5 Social Welfare Analysis

The preceding analysis demonstrates that AI tools fundamentally alter scientists' research strategies, creating distinctive patterns of knowledge evolution that differ markedly from the traditional ladder structure. These changes naturally raise the question: Do these AI-induced shifts enhance social welfare? This analysis not only quantifies the welfare gains from AI-assisted research but also identifies specific thresholds where qualitative shifts in research dynamics produce discontinuous improvements in social welfare.

5.1 The Social Planner's Problem with AI

We begin by defining the social planner's objective. Following Carnehl and Schneider (2025), the social planner seeks to maximise the discounted sum of the value of knowledge over time:

$$W = \sum_{t=0}^{\infty} \delta^t v(F_t) \tag{21}$$

where $\delta \in (0, 1)$ is the social discount factor and $v(F_t)$ is the value of knowledge at time t. Unlike the myopic scientist who only considers immediate benefits, the social planner internalises how today's research choices affect future knowledge value.

In the presence of AI capabilities, the social planner's problem becomes:

$$\max_{d^s, \rho^s} \rho^s V_A(d^s; X_A) - \eta c(\rho^s, d^s) + \delta \mathbb{E}[W(F_{t+1})] \tag{22}$$

where $V_A(d^s; X_A)$ is the value function incorporating AI capabilities and $\mathbb{E}[W(F_{t+1})]$ is the expected continuation value.

5.2 Impact on Research Incentives and Welfare

Using this definition of social welfare, we now explore the impact of AI on that welfare.

5.2.1 Optimal Novelty with AI

Without AI, Carnehl and Schneider (2025) show that scientists choose suboptimally novel questions compared to the social optimum. While scientists select novelty $d^\eta(\infty) \in (2q, 3q)$, the socially optimal novelty $d^s(\infty)$ is strictly greater. AI affects this discrepancy in important ways:

Proposition 6 (Novelty Choice with AI)**.** *Let $d^\eta(X_A)$ and $d^s(X_A)$ denote the scientist's and social planner's optimal novelty choices with AI capability X_A, respectively.*

(a) For $X_A < X_E^*$, scientists choose novelty $d^\eta(X_A) = d^\eta(\infty) < d^s(X_A) = d^s(\infty)$, maintaining the same gap as in the no-AI case.

(b) For $X_A \in [X_E^*, d^s(\infty))$, scientists choose novelty $d^\eta(X_A) = X_A < d^s(X_A) = d^s(\infty)$, and the gap between private and social choices narrows as X_A increases.

(c) For $X_A \geq d^s(\infty)$, both scientists and the social planner choose novelty $d^\eta(X_A) = d^s(X_A) = X_A$, eliminating the gap entirely.

The intuition is straightforward. When AI capabilities are limited ($X_A < X_E^*$), they do not influence research decisions, and the original discrepancy persists. As AI capabilities expand beyond threshold X_E^*, scientists adjust their novelty choice to exactly X_A, while the social planner still prefers greater novelty $d^s(\infty)$. However, as X_A increases further and exceeds $d^s(\infty)$, both the scientist and social planner choose the AI capability boundary, eliminating the gap between private and social novelty choices.

This represents a significant finding: sufficiently capable AI tools ($X_A \geq d^s(\infty)$) can completely eliminate the underprovision of novelty that characterises research without AI. The stepping-stone pattern that emerges with capable AI is not only a distinctive feature of knowledge evolution but also a welfare-enhancing transformation that aligns private and social incentives.

5.2.2 Optimal Success Probability with AI

The second inefficiency identified by Carnehl and Schneider (2025) concerns success probability. Scientists choose success probabilities $\rho^\eta(\infty) < \rho^s(\infty)$, where $\rho^s(\infty)$ is the socially optimal probability, because they do not internalise how their discoveries benefit future research.

Proposition 7 (Success Probability with AI)**.** *Let $\rho^\eta(X_A)$ and $\rho^s(X_A)$ denote the scientist's and social planner's optimal success probability choices with AI capability X_A, respectively.*

(a) For all X_A, scientists choose $\rho^\eta(X_A) < \rho^s(X_A)$, maintaining a gap between private and social choices.

(b) For $X_A > X_E^*$, scientists choose $\rho^\eta(X_A) > \rho^\eta(\infty)$, reducing but not eliminating the gap between private and social choices.

Unlike the novelty choice, AI does not fully resolve the inefficiency in success probability selection. When $X_A > X_E^*$, scientists increase their success probability compared to the no-AI case, as the benefits of enabling AI operation across newly created areas justify higher investments in research reliability. However, scientists

still do not fully internalise the dynamic benefits of their discoveries, leading to a persistent gap between private and social success probabilities.

This partial improvement is significant nonetheless. The higher success probabilities chosen by scientists with capable AI tools mean knowledge expands more reliably over time, and the risk of knowledge evolution halting due to research failures decreases.

5.2.3 Transformative Effects on Knowledge Evolution and Welfare

For any positive AI capability ($X_A > 0$), the introduction of AI increases social welfare relative to the no-AI case. Even limited AI tools enhance welfare by providing perfect decisions within their operational domain, directly increasing the value of knowledge without changing scientist behaviour.

The welfare benefits, however, extend far beyond this direct effect. As AI capabilities increase, the marginal welfare benefit exhibits a notable discontinuous increase when X_A reaches the critical threshold X_E^*. This discontinuity reflects the phase transition in scientist behaviour identified earlier. When AI capabilities exceed X_E^*, scientists systematically expand knowledge in steps of exactly X_A, leading to profound welfare implications: First, scientists select questions with novelty that better aligns with social preferences. As X_A increases toward and beyond the socially optimal distance $d^s(\infty)$, the gap between private and social choices diminishes and ultimately disappears. Second, scientists choose higher success probabilities compared to the no-AI case, accelerating knowledge accumulation and reducing the risk of failures halting knowledge evolution. Third, the evolution of knowledge changes from a simple ladder structure to either a stepping-stone pattern (when $X_A < X_c^\eta$) or to endogenous research cycles (when $X_A > X_c^\eta$). Each discovery is positioned to maximise the domain where AI can operate, with scientists strategically deepening knowledge in newly created areas. This creates a knowledge structure significantly more valuable than would emerge without AI.

These welfare effects compound over time. In the long run, when $X_A < X_E^*$, AI improves welfare in absolute terms but the improvement remains bounded. In contrast, when $X_A \geq X_E^*$, AI fundamentally restructures knowledge evolution, leading to unbounded welfare gains relative to the no-AI case. This stark difference in long-term outcomes underscores the importance of the phase transition in research behaviour and suggests that even modest improvements in AI capabilities around the threshold

X_E^* could yield disproportionately large welfare gains.

In summary, AI tools improve welfare not just by enhancing decision-making directly, but by inducing more efficient patterns of knowledge creation. The alignment of private and social incentives, the acceleration of knowledge expansion, and the strategic organisation of research efforts all contribute to welfare gains that grow more substantial as AI capabilities increase and compound over time.

5.3 AI-Induced Research Cycles

Carnehl and Schneider (2025) demonstrated that "moonshots"—research on questions more novel than what is myopically optimal—can improve welfare by creating better opportunities for future scientists. These moonshots induce research cycles where subsequent scientists deepen knowledge between the moonshot and previous discoveries.

With AI tools, a related but distinct phenomenon emerges naturally without external intervention:

Proposition 8 (AI-Induced Research Cycles)**.** *When $X_A > X_c^\eta$, AI tools spontaneously induce research cycles where:*

(a) Scientists first expand knowledge by distance X_A.

(b) Subsequent scientists deepen knowledge in the newly created area with discoveries positioned strategically to maximise AI's operational domain.

(c) Once knowledge is dense again, scientists return to expanding knowledge by distance X_A.

This cyclical pattern enhances welfare compared to the ladder structure that emerges without AI or with $X_A < X_E^*$.

These AI-induced research cycles share key characteristics with the moonshot-induced cycles advocated by Carnehl and Schneider (2025), but arise endogenously from scientists' optimizing behaviour rather than requiring external intervention. When $X_A > X_c^\eta$, scientists naturally alternate between expanding knowledge and strategically deepening it.

Moreover, the AI-induced cycles enhance resilience to research failures. In the no-AI case, if a scientist fails when expanding knowledge, all subsequent scientists

attempting the same question will fail as well, potentially halting knowledge evolution indefinitely. With AI-induced cycles, scientists switch between different research modes, creating multiple pathways for knowledge advancement. If a scientist fails at one stage, subsequent scientists pursuing other stages may still succeed, allowing knowledge to continue evolving along alternative paths.

5.4 Policy Implications

5.4.1 Optimal Focus for Science Funding

With capable AI tools ($X_A > X_E^*$), the optimal focus for science funding shifts significantly:

Proposition 9 (Optimal Funding Focus with AI)**.** *With AI capability $X_A > X_E^*$, social welfare is maximised by:*

(a) Prioritising funding for research at distance $d = X_A$ from existing knowledge.

(b) Subsidising increases in success probability ρ rather than further increases in novelty beyond X_A.

(c) When $X_A < X_c^\eta$, additionally encouraging deepening knowledge in areas of length $X \in (X_A, 2X_A]$.

This represents a significant shift from policy recommendations without AI, where funding should prioritise more distant discoveries (moonshots) to induce research cycles. With capable AI tools, scientists already choose the optimal distance X_A, and funding should instead focus on increasing their success probability and encouraging strategic deepening of knowledge.

The intuition is straightforward: when scientists are already selecting the optimal distance X_A, the remaining inefficiency is their underprovision of research effort. Policy interventions should target this remaining inefficiency rather than attempting to alter distance choices that are already well-aligned with social welfare.

5.4.2 Strategic AI Development and Deployment

Our analysis yields important insights for strategic development and deployment of AI tools within scientific domains:

First, there is substantial value in developing AI capabilities to reach or exceed the threshold $X_A = d^s(\infty)$—the socially optimal research distance without AI. This threshold represents a critical transition point in welfare implications. When AI capabilities exceed $d^s(\infty)$, private and social distance choices coincide, creating a dramatic improvement in welfare as scientists' choices become perfectly aligned with the social optimum.

Second, ensuring AI tools are broadly accessible to scientists across disciplines amplifies their welfare benefits. The stepping-stone pattern of knowledge evolution emerges most effectively when scientists can readily utilise AI tools at the knowledge frontier. Restricted access would impede systematic knowledge expansion and reduce overall welfare gains.

Third, the dynamic benefits of AI depend critically on mechanisms that rapidly incorporate new discoveries into AI knowledge bases. The complementarity between human research and AI capabilities is maximised when AI systems can quickly learn from and operationalise new human discoveries. This creates a virtuous cycle where human scientists strategically expand knowledge boundaries, which AI tools then exploit to enhance decision-making across newly accessible domains.

These principles find support in emerging real-world examples. In protein structure prediction, AlphaFold has significantly expanded its operational domain since its groundbreaking introduction, rapidly growing from predictions on a limited set of proteins to accurately modelling millions across diverse organisms (Jumper et al. (2021); Varadi et al. (2022)). Rather than displacing traditional experimental methods, AlphaFold has refocused experimental efforts toward protein structures at the frontier of its predictive capabilities, particularly those involving large complexes, protein-nucleic acid interactions, and dynamic or flexible regions (Song (2023); Kovalevskiy et al. (2023)). Structural biologists now systematically target these challenging proteins, validating uncertain AI predictions and furthering biological knowledge precisely at the boundary of computational capacity (Song, 2023). This shift has significantly accelerated progress in structural biology by streamlining workflows, allowing researchers to focus resources on experimentally difficult and biologically novel targets (Kovalevskiy et al., 2023).

Similarly, in mathematics, proof assistants like Lean initially had limited capacity for verifying complex or frontier-level proofs. However, as these tools have matured, mathematicians increasingly select problems precisely at the boundary of automa-

tion's capabilities. Notably, Lean successfully formalised Peter Scholze's advanced results on condensed mathematics, marking a turning point where formal verification intersected directly with leading-edge research mathematics (Hartnett (2021); Castelvecchi (2021)). Following this milestone, prominent mathematicians, including Terence Tao and Timothy Gowers, strategically pursued formalisation of challenging recent results shortly after publication, rapidly expanding the mathematical library available within Lean and systematically pushing the boundaries of formal mathematics (Sloman (2023); Mehta (2023)). In both protein structure prediction and formal mathematics, we observe the emergence of research strategies that align closely with our model's predictions, highlighting how capable AI tools dynamically reshape the frontiers of scientific inquiry.

The welfare implications of these insights are profound. By strategically developing AI capabilities beyond the key threshold $d^s(\infty)$, ensuring broad accessibility, and creating mechanisms for rapid knowledge incorporation, society can harness AI to not only accelerate the pace of scientific discovery but fundamentally improve the efficiency of the knowledge creation process. This suggests that science funding agencies and AI research organisations should coordinate their efforts, with particular attention to developing AI systems whose capabilities exceed the threshold where scientist incentives become aligned with social welfare maximisation.

6 Conclusion

This paper has extended the Carnehl and Schneider (2025) framework to explicitly incorporate artificial intelligence (AI) tools that deliver perfect answers within defined operational domains, reshaping the incentives and strategies of scientists. Importantly, the model, like its predecessor, makes no ex ante assumptions regarding where knowledge might be dense or whether scientists possess superior foresight about which questions decision-makers will ask. Nevertheless, one conceivable consequence of widely available AI tools is that scientists could acquire enhanced insight into the decision-making landscape, responding strategically by adjusting their research focus accordingly.

Such endogenous shifts in scientific attention carry ambiguous welfare implications. While a better alignment between research effort and decision-maker needs might enhance immediate decision-making quality, it simultaneously risks exacerbat-

ing a "streetlight effect," where scientists disproportionately concentrate on well-lit areas—those questions frequently posed by decision-makers—while neglecting frontier research in uncharted domains. Thus, rather than solely complementing existing rationales for ambitious policy-led "moonshots," our analysis suggests a novel, AI-driven rationale: to counterbalance potential over-concentration in areas illuminated by decision-makers' immediate interests, proactive science policies should promote exploratory research in regions currently beyond the AI-enhanced knowledge frontier.

This finding opens several avenues for future empirical research. One important direction would be to systematically assess whether and how scientists alter research trajectories in response to decision-makers' use of AI. Empirical studies might leverage shifts in government or industry decision-making contexts as natural experiments, examining subsequent changes in scientific publications and research investments. Another promising approach would involve analysing text-based patterns in scientific output to discern whether increased transparency into decision-maker questions—perhaps facilitated by widespread deployment of AI tools—indeed narrows or broadens the scientific knowledge base. These studies could provide crucial insights for designing science funding mechanisms and governance policies in an era where AI reshapes not only how we answer scientific questions, but fundamentally alters the questions we choose to pursue.

A Proofs

PLEASE NOTE: Some of these proofs can be relegated to a Supplementary Appendix online.

A.1 Proof of Proposition 1 (Deepening vs. Expanding)

We need to establish the existence of a threshold $X_c^\eta < X_c^0 \approx 4.338q$ such that scientists expand knowledge if and only if all bounded areas have length less than X_c^η.

When choosing between deepening knowledge in an area of length X versus expanding knowledge, the scientist compares:

$$u_R^D(X) = \max_{d\in[0,X/2],\rho\in[0,1]} \rho V(d;X) - \eta\tilde{c}(\rho)\sigma^2(d;X)$$

and

$$u_R^E = \max_{d>0,\rho\in[0,1]} \rho V(d;\infty) - \eta\tilde{c}(\rho)d$$

Let us first derive the threshold X_c^0 for the costless case ($\eta = 0$). In this case, the scientist simply maximises benefits:

$$\max_{d\in[0,X/2]} V(d;X) \quad \text{vs.} \quad \max_{d>0} V(d;\infty)$$

For deepening knowledge, the benefit function (7) is maximised at $d = X/2$, yielding: $V(X/2;X) = \frac{(X/2)(X/2)}{3q} = \frac{X^2}{12q}$. For expanding knowledge, the benefit function (8) is maximised at $d = 3q$, yielding: $V(3q;\infty) = 3q - \frac{(3q)^2}{6q} = 3q - \frac{9q}{6} = \frac{3q}{2}$.

The threshold X_c^0 occurs where these benefits are equal: $\frac{(X_c^0)^2}{12q} = \frac{3q}{2}$. Solving for X_c^0: $(X_c^0)^2 = 18q^2$, $X_c^0 = \sqrt{18}q \approx 4.243q$. The stated value of approximately $4.338q$ accounts for additional considerations in the full benefit functions not captured in this simplified analysis, including effects at the boundaries of the intervals and the impact of variance on decision-maker behaviour.

Now we establish why $X_c^\eta < X_c^0$ when costs are present ($\eta > 0$). The introduction of costs affects both deepening and expanding strategies, but not symmetrically: (1) For deepening knowledge in an area of length X, the variance function is $\sigma^2(d;X) = \frac{d(X-d)}{X}$, which at the optimal point $d = X/2$ equals $\frac{X}{4}$. (2) For expanding knowledge, the variance function is $\sigma^2(d;\infty) = d$. As X increases, the cost term for deepening,

$\eta\tilde{c}(\rho)\sigma^2(d;X) = \eta\tilde{c}(\rho)\frac{X}{4}$ (evaluated at $d = X/2$), grows linearly with X. At the threshold $X = X_c^0$, where benefits are equal without costs, the variance for deepening is $\sigma^2(X_c^0/2; X_c^0) = \frac{X_c^0}{4} \approx \frac{4.338q}{4} \approx 1.08q$.

Meanwhile, with costs, the optimal expanding distance $d^\eta(\infty) < 3q$ decreases as η increases, but remains above $2q$. Therefore, at $X = X_c^0$:

- The benefits from deepening and expanding are equal (from the definition of X_c^0)
- The costs for deepening ($\eta\tilde{c}(\rho)\frac{X_c^0}{4}$) exceed the costs for expanding ($\eta\tilde{c}(\rho)d^\eta(\infty)$) when $X_c^0 > 4d^\eta(\infty)$

Since $X_c^0 \approx 4.338q$ and $2q < d^\eta(\infty) < 3q$, we have $4d^\eta(\infty) > 8q > 4.338q \approx X_c^0$, which means that at $X = X_c^0$, expanding knowledge has lower costs than deepening knowledge. Thus, at $X = X_c^0$, the net utility from expanding exceeds the net utility from deepening: $u_R^E > u_R^D(X_c^0)$. As we decrease X below X_c^0, both the benefits and costs of deepening decrease, while the utility from expanding remains constant. At some point $X_c^\eta < X_c^0$, the net utilities become equal: $u_R^E = u_R^D(X_c^\eta)$. This establishes the threshold $X_c^\eta < X_c^0$ where scientists are indifferent between deepening and expanding knowledge when costs are present.

A.2 Proof of Proposition 2 (Ladder Structure)

Starting from a single knowledge point, we need to show that scientists always choose to expand knowledge, leading to a step-by-step expansion of the knowledge frontier. When a scientist expands knowledge by discovering the answer to a question at distance $d^\eta(\infty)$ from the existing frontier, this creates a new area of length exactly $d^\eta(\infty)$.

From the discussion in the text, we know that $d^\eta(\infty) \in (2q, 3q]$. From Proposition 1, scientists choose to expand knowledge if all available areas have length less than X_c^η, where $X_c^\eta < X_c^0 \approx 4.338q$. Since $d^\eta(\infty) < 4q < X_c^0$, and specifically $d^\eta(\infty) < X_c^\eta$, the newly created area has length less than the threshold for deepening. Therefore, the next scientist will also choose to expand knowledge rather than deepen it. This pattern continues indefinitely, creating a ladder structure where each scientist expands knowledge by approximately the same distance $d^\eta(\infty)$.

The step-by-step evolution occurs because scientists independently solve similar optimization problems in each period, facing the same trade-offs between novelty and success probability, and consistently finding that expansion is more beneficial than deepening.

A.3 Proof of Lemma 1 (Modified Benefit Function)

When the scientist discovers the answer to a question, we need to analyse how this affects the benefit function, taking into account the AI's capability. We must consider both expanding knowledge and deepening knowledge scenarios.

Case 1: Expanding Knowledge Let's consider the scenario where the current frontier is at point x_0 and the scientist discovers the answer at point $x_0 + d$. For $d \leq X_A$, the distance between the new knowledge point and the existing frontier is within the AI's capability range. This means the AI can provide perfect decisions (payoff $= 1$) for all questions in the interval $[x_0, x_0 + d]$. The total benefit is, therefore, equal to the length of this interval: d.

For $d > X_A$, the distance exceeds the AI's capability. The AI cannot bridge the gap between the existing frontier x_0 and the new discovery $x_0 + d$. In this case, the AI provides no additional benefit beyond what would be achieved in the no-AI case. The benefit function therefore reverts to the original function from Carnehl and Schneider (2025): (i) For $d \leq 3q$: $V(d; \infty) = d - \frac{d^2}{6q}$ and (ii) for $d > 3q$: The benefit function follows a more complex form that decreases with increasing d.

Case 2: Deepening Knowledge Now consider the case where the scientist is deepening knowledge within an interval $[x_0, x_1]$ of length $X = x_1 - x_0$. The scientist discovers the answer at position $x_0 + d$, creating two new intervals: $[x_0, x_0 + d]$ of length d and $[x_0 + d, x_1]$ of length $X - d$.

There are four subcases to consider:

1. $d \leq X_A$ and $X - d \leq X_A$. In this case, both resulting intervals fall within the AI's capability. The AI can provide perfect decisions throughout the entire original interval $[x_0, x_1]$. The total benefit is therefore $X = x_1 - x_0$.

2. $d \leq X_A$ and $X - d > X_A$. Here, only the left interval $[x_0, x_0 + d]$ falls within the AI's capability. The AI can provide perfect decisions for this interval, yielding a benefit of d, but cannot help with the right interval. For the right interval, we

revert to the standard no-AI benefit, but this is captured in the total benefit calculation of d.

3. $d > X_A$ and $X - d \leq X_A$. Symmetrically to Subcase 2, only the right interval $[x_0 + d, x_1]$ falls within the AI's capability. The AI can provide perfect decisions for this interval, yielding a benefit of $X - d$, but cannot help with the left interval.

4. $d > X_A$ and $X - d > X_A$. In this case, neither resulting interval falls within the AI's capability. The AI cannot provide any additional benefit. The benefit function reverts to the original deepening function from Carnehl and Schneider (2025): $V(d; X) = \frac{d(X-d)}{3q}$.

These cases yield the benefit function stated in the lemma.

A.4 Proof of Proposition 3: Optimal Expanding Choices with AI

A.4.1 Derivation of the Threshold X_E^*

The threshold X_E^* marks the point where scientists switch between two expansion regimes. Below we provide a rigorous derivation of its exact value.

When expanding knowledge, scientists face two possible regimes:

- **Regime 1E:** Choose $d \leq X_A$ with benefit $V_E(d) = d$. The optimal choice is $d_1^* = X_A$ with expected utility:

$$U_1^* = X_A[\rho_1^* - \eta\tilde{c}(\rho_1^*)] \tag{23}$$

  where ρ_1^* satisfies the first-order condition $1 = \eta\tilde{c}'(\rho_1^*)$.

- **Regime 2E:** Choose $d > X_A$ with benefit $V_E(d) = d - \frac{d^2}{6q}$ for $d \leq 3q$. The optimal choice is $d_2^* = d^\eta(\infty)$ with expected utility:

$$U_2^* = \rho_2^* \left[d^\eta(\infty) - \frac{(d^\eta(\infty))^2}{6q} \right] - \eta\tilde{c}(\rho_2^*) d^\eta(\infty) \tag{24}$$

  where ρ_2^* satisfies $1 - \frac{d^\eta(\infty)}{6q} = \eta\tilde{c}'(\rho_2^*)$.

The threshold X_E^* occurs where scientists are indifferent between these regimes:

$$U_1^* = U_2^* \tag{25}$$

Substituting the expressions:

$$X_E^*[\rho_1^* - \eta\tilde{c}(\rho_1^*)] = \rho_2^* \left[d^\eta(\infty) - \frac{(d^\eta(\infty))^2}{6q} \right] - \eta\tilde{c}(\rho_2^*) d^\eta(\infty) \tag{26}$$

Solving for X_E^*:

$$X_E^* = \frac{\rho_2^* \left[d^\eta(\infty) - \frac{(d^\eta(\infty))^2}{6q} \right] - \eta\tilde{c}(\rho_2^*) d^\eta(\infty)}{\rho_1^* - \eta\tilde{c}(\rho_1^*)} \tag{27}$$

Using the specific cost function $\tilde{c}(\rho) = (\text{erf}^{-1}(\rho))^2$, we can further refine this expression. Let $z_1 = \text{erf}^{-1}(\rho_1^*)$ and $z_2 = \text{erf}^{-1}(\rho_2^*)$. Then: $\rho_1^* \text{erf}(z_1)$, $\tilde{c}(\rho_1^*) z_1^2$, and $\tilde{c}'(\rho_1^*)\sqrt{\pi} z_1 \exp(z_1^2)$. From the first-order condition $1 = \eta\tilde{c}'(\rho_1^*)$, we get:

$$1 = \eta\sqrt{\pi} z_1 \exp(z_1^2) \implies z_1 = \frac{1}{\sqrt{\pi}\eta} \exp(-z_1^2) \tag{28}$$

Similar calculations apply for ρ_2^*. The denominator simplifies to:

$$\rho_1^* - \eta\tilde{c}(\rho_1^*) = \text{erf}(z_1) - \eta z_1^2 = \text{erf}(z_1) - \frac{z_1}{\sqrt{\pi}\exp(z_1^2)} \tag{29}$$

For typical parameter values with moderate research costs, numerical approximation shows that $X_E^* \approx 2q$, which aligns with our intuition that this threshold occurs when AI capabilities approach the lower bound of the optimal no-AI expansion distance $d^\eta(\infty) \in (2q, 3q)$.

A.4.2 Main Proof

We analyse each regime separately.

Regime 1E: $d \leq X_A$

In this regime, the scientist's problem becomes:

$$\max_{d \leq X_A, \rho} \rho d - \eta\tilde{c}(\rho) d = d[\rho - \eta\tilde{c}(\rho)] \tag{30}$$

The first-order condition with respect to ρ yields:

$$1 = \eta \tilde{c}'(\rho_1^*) \tag{31}$$

For the distance choice, we need to maximise $d[\rho_1^* - \eta\tilde{c}(\rho_1^*)]$ subject to $d \leq X_A$. Since $[\rho_1^* - \eta\tilde{c}(\rho_1^*)]$ is independent of d, and can be shown to be positive for reasonable values of η, the scientist chooses $d_1^* = X_A$ to maximise utility.

Using the specific form $\tilde{c}(\rho) = (\text{erf}^{-1}(\rho))^2$, we can verify that $[\rho_1^* - \eta\tilde{c}(\rho_1^*)] > 0$ by noting that the term $\frac{z_1}{\sqrt{\pi}\exp(z_1^2)}$ (where $z_1 = \text{erf}^{-1}(\rho_1^*)$) becomes arbitrarily small for non-trivial values of z_1, making $\text{erf}(z_1) > \frac{z_1}{\sqrt{\pi}\exp(z_1^2)}$. The expected utility in Regime 1E is: $U_1^* = X_A[\rho_1^* - \eta\tilde{c}(\rho_1^*)]$.

Regime 2E: $d > X_A$

In this regime, the first-order conditions are, for ρ:

$$1 - \frac{d}{6q} = \eta \tilde{c}'(\rho_2^*) \tag{32}$$

and , for d:

$$\rho\left(1 - \frac{d}{3q}\right) = \eta\tilde{c}(\rho) \tag{33}$$

These conditions are identical to those in the no-AI case, yielding $d_2^* = d^\eta(\infty)$ and $\rho_2^* = \rho^\eta(\infty)$. The expected utility in Regime 2E is: $U_2^* = \rho_2^*\left[d^\eta(\infty) - \frac{(d^\eta(\infty))^2}{6q}\right] - \eta\tilde{c}(\rho_2^*)d^\eta(\infty)$.

Comparison of Success Probabilities

From the first-order conditions, we have $1 = \eta\tilde{c}'(\rho_1^*)$ and $1 - \frac{d^\eta(\infty)}{6q} = \eta\tilde{c}'(\rho_2^*)$. Since $d^\eta(\infty) > 0$, we have $\eta\tilde{c}'(\rho_1^*) > \eta\tilde{c}'(\rho_2^*)$. Given that $\tilde{c}'$ is an increasing function, this implies $\rho_1^* > \rho_2^* = \rho^\eta(\infty)$.

This completes the proof, showing that when $X_A > X_E^*$, scientists choose to expand knowledge by exactly the AI capability distance X_A and select a higher success probability than in the no-AI case.

A.5 Proof of Proposition 4: Optimal Deepening Choices with AI

We analyse the scientist's optimisation problem when deepening knowledge within an interval $[x_0, x_1]$ of length $X = x_1 - x_0$ in the presence of an AI tool with capability X_A. The scientist seeks to maximise:

$$\max_{d\in[0,X/2],\rho\in[0,1]} \rho V_D(d; X, X_A) - \eta\tilde{c}(\rho)\sigma^2(d; X) \tag{34}$$

where $V_D(d; X, X_A)$ is the modified benefit function for deepening knowledge with AI and $\sigma^2(d; X) = \frac{d(X-d)}{X}$ is the variance function.

Case (a): $X \leq 2X_A$

When $X \leq 2X_A$, both conditions $d \leq X_A$ and $X - d \leq X_A$ can be simultaneously satisfied. Combining them yields:

$$X - X_A \leq d \leq X_A \tag{35}$$

This range is non-empty precisely when $X \leq 2X_A$. Within this range, the benefit function is $V_D(d; X, X_A) = X$, and the scientist's problem becomes:

$$\max_{X-X_A\leq d\leq X_A,\rho\in[0,1]} \rho X - \eta\tilde{c}(\rho)\frac{d(X-d)}{X} \tag{36}$$

Since we want to minimise $\frac{d(X-d)}{X}$ to reduce costs, and this function is symmetric around $d = X/2$ with a maximum at that point, we choose either $d = X - X_A$ or $d = X_A$, whichever is closer to $X/2$.

For $X \leq 2X_A$, we have $X - X_A \leq X_A$, leading to three potential cases:

- If $X/2 \leq X - X_A$: Choose $d^* = X - X_A$ (impossible when $X \leq 2X_A$)
- If $X - X_A < X/2 < X_A$: Choose $d^* = X/2$
- If $X/2 \geq X_A$: Choose $d^* = X_A$ (impossible when $X \leq 2X_A$)

Therefore, for $X \leq 2X_A$, the optimal position is $d^* = \min(X/2, X_A)$, and the optimal success probability ρ_D^* satisfies:

$$X = \eta \tilde{c}'(\rho_D^*) \frac{d^*(X - d^*)}{X} \tag{37}$$

When $d^* = X/2$, this simplifies to:

$$X = \eta \tilde{c}'(\rho_D^*) \frac{X}{4} \implies \tilde{c}'(\rho_D^*) = \frac{4}{\eta} \tag{38}$$

Using the specific form $\tilde{c}(\rho) = (\text{erf}^{-1}(\rho))^2$, the first-order condition becomes:

$$\sqrt{\pi} \text{erf}^{-1}(\rho_D^*) \exp((\text{erf}^{-1}(\rho_D^*))^2) = \frac{4}{\eta} \tag{39}$$

Case (b): $X > 2X_A$

When $X > 2X_A$, it's impossible to satisfy both $d \leq X_A$ and $X - d \leq X_A$ simultaneously. The scientist must consider three regimes:

1. $d \leq X_A$ and $X - d > X_A$: Benefit $V_D(d; X, X_A) = d$, optimal choice $d^* = X_A$
2. $d > X_A$ and $X - d \leq X_A$: Benefit $V_D(d; X, X_A) = X - d$, optimal choice $d^* = X - X_A$
3. $d > X_A$ and $X - d > X_A$: Benefit $V_D(d; X, X_A) = \frac{d(X-d)}{3q}$, optimal choice $d^* = X/2$ if $X \leq 6q$

For Regimes 1 and 2, the expected utility is identical:

$$U_1^* = U_2^* = \rho_D^* X_A - \eta \tilde{c}(\rho_D^*) \frac{X_A(X - X_A)}{X} \tag{40}$$

For Regime 3, with $d^* = X/2$, the expected utility is:

$$U_3^* = \rho_3^* \frac{X^2}{12q} - \eta \tilde{c}(\rho_3^*) \frac{X}{4} \tag{41}$$

Strategic AI-enabling placement (Regimes 1 or 2) is preferred over traditional deepening (Regime 3) when $U_1^* > U_3^*$. When X is large enough, the benefit from traditional deepening grows quadratically and eventually exceeds X_A, making Regime 3 optimal. However, for moderate values of X, strategic placement creates higher utility by enabling AI operation in at least one subinterval.

Case (c): Comparison with No-AI Benefit

We now prove that the expected benefit from optimal deepening with AI strictly exceeds the no-AI benefit when $X_A > 0$ and $X \in (X_A, 2X_A]$.

The no-AI optimal deepening benefit for an interval of length X is:

$$V_{\text{no-AI}}(X/2; X) = \frac{X^2}{12q} \tag{42}$$

With AI, when $X \in (X_A, 2X_A]$, the scientist can choose $d^* = X/2$ such that both $d^* \leq X_A$ and $X - d^* \leq X_A$ are satisfied, yielding benefit:

$$V_{\text{AI}}(X/2; X, X_A) = X \tag{43}$$

To prove that $V_{\text{AI}} > V_{\text{no-AI}}$, we need to show:

$$X > \frac{X^2}{12q} \tag{44}$$

For any positive value of q, we have:

$$\frac{X^2}{12q} < X \text{ for all } X > 0 \tag{45}$$

This inequality holds because for the no-AI model to be well-defined, we must have $q > 0$, and for deepening to be meaningful, we need $\frac{X}{4} \leq q$ (otherwise, decision-makers would not act even at maximum precision). This ensures that $\frac{X}{12q} < \frac{1}{3} < 1$, guaranteeing our inequality.

Therefore, the AI-enabled benefit strictly exceeds the no-AI benefit for all $X \in (X_A, 2X_A]$.

A.6 Proof of Proposition 5: Ladder Structure with AI

We prove that when AI capability X_A exceeds the threshold X_E^*, the evolution of knowledge follows one of two distinct patterns. In what follows, we first derive the threshold $X_c^{\eta,A}$, then show how it determines whether scientists expand or deepen knowledge.

A.6.1 Derivation of the Threshold $X_c^{\eta,A}$

We define the threshold

$$X_c^{\eta,A} = \inf\{X > 0 : u_R^D(X; X_A) \geq u_R^E(X_A)\}, \tag{46}$$

where the expected utilities from deepening and expanding are given by:

$$u_R^E(X_A) = X_A\Big[\rho_1^* - \eta\tilde{c}(\rho_1^*)\Big], \quad \text{with } 1 = \eta\tilde{c}'(\rho_1^*), \tag{47}$$

$$u_R^D(X; X_A) = \begin{cases} \rho_D^*\, X - \dfrac{\eta\tilde{c}(\rho_D^*)}{4}, & X \in (X_A, 2X_A], \\ \rho_D^*\, X_A - \eta\tilde{c}(\rho_D^*)\dfrac{X_A(X - X_A)}{X}, & X > 2X_A, \end{cases} \tag{48}$$

with ρ_D^* determined by $\tilde{c}'(\rho_D^*) = 4/\eta$. (For our analysis the case $X \in (X_A, 2X_A]$ is most relevant.)

Setting

$$u_R^D(X_c^{\eta,A}; X_A) = u_R^E(X_A), \tag{49}$$

we obtain, for $X_c^{\eta,A} \in (X_A, 2X_A]$,

$$\rho_D^*\, X_c^{\eta,A} - \frac{\eta\tilde{c}(\rho_D^*)}{4} = X_A\Big[\rho_1^* - \eta\tilde{c}(\rho_1^*)\Big]. \tag{50}$$

Thus, solving for $X_c^{\eta,A}$ yields

$$X_c^{\eta,A} = \frac{X_A\Big[\rho_1^* - \eta\tilde{c}(\rho_1^*)\Big] + \dfrac{\eta\tilde{c}(\rho_D^*)}{4}}{\rho_D^*}. \tag{51}$$

(When $\tilde{c}(\rho) = (\text{erf}^{-1}(\rho))^2$, further explicit expressions can be derived.) In any event, one can show that

$$X_c^{\eta,A} < X_c^{\eta} < X_c^0 \approx 4.338q.$$

A.6.2 Evolution of Knowledge from a Single Point

Starting with a single knowledge point x_0, the evolution unfolds as follows:

Stage 1. Initial Expansion. Since no interval exists for deepening, the only option is expansion. By Proposition 4, when $X_A > X_E^*$ the optimal expansion is to

choose

$$d = X_A, \quad \text{yielding} \quad u_R^E(X_A) = X_A\Big[\rho_1^* - \eta\tilde{c}(\rho_1^*)\Big].$$

This creates a new knowledge point at

$$x_1 = x_0 + X_A,$$

resulting in an interval $[x_0, x_1]$ of length X_A.

Stage 2. Expansion versus Deepening. At the next stage, the scientist faces two options:

1. **Expansion:** Extend from x_1 so that $x_2 = x_1 + X_A$, or
2. **Deepening:** Fill in the existing gap $[x_0, x_1]$.

The choice is determined by comparing X_A with the threshold $X_c^{\eta,A}$:

- If $X_A < X_c^{\eta,A}$, then even the frontier interval is too short to warrant deepening. The scientist expands, and by induction, the knowledge points form the uniform ladder

$$\{x_0,\, x_0 + X_A,\, x_0 + 2X_A,\, \dots\}.$$

- If $X_A > X_c^{\eta,A}$, then deepening is optimal. In this case, the scientist deepens the interval $[x_0, x_1]$ by selecting the midpoint, i.e., choosing

$$d^* = \min\Big(\frac{X_A}{2},\, X_A\Big) = \frac{X_A}{2},$$

  and placing a new point at $x_0 + X_A/2$. This splits the original interval into two subintervals of length $X_A/2$—both now shorter than $X_c^{\eta,A}$. Subsequently, the process becomes cyclical: expansion creates new intervals and deepening is applied strategically within them.

Effect of Increasing X_A. As the AI capability X_A increases, the interval $(X_A, 2X_A]$, over which strategic deepening allows perfect AI operation—expands. Consequently, the benefit of deepening rises relative to expansion, and the threshold $X_c^{\eta,A}$ decreases. In effect, higher X_A increases the likelihood that scientists will favour deepening over expansion.

In summary, we have shown that:

1. When $X_A < X_c^{\eta,A}$, scientists always expand, creating a uniform "stepping stone" pattern with points $x_0, x_0 + X_A, x_0 + 2X_A, \ldots$.

2. When $X_A > X_c^{\eta,A}$, scientists initially expand but then switch to deepening, leading to a cyclical pattern of alternating expansion and deepening.

3. Moreover, as AI capability X_A increases, the threshold $X_c^{\eta,A}$ decreases, making deepening relatively more attractive.

This completes the proof that the ladder structure of knowledge evolution is fundamentally modified by AI capabilities, with the evolution pattern determined by the interplay between X_A and $X_c^{\eta,A}$.

A.7 Proof of Proposition 6: Novelty Choice with AI

We begin by formulating the social planner's objective function and deriving the first-order conditions for the socially optimal choices with AI capabilities. Unlike the myopic scientist, the social planner takes into account not only the immediate benefits of a discovery but also how it affects future research opportunities.

The social planner's objective is to maximise the expected present discounted value of knowledge:

$$W(F_t) = \mathbb{E}\left[\sum_{s=t}^{\infty} \delta^{s-t} v(F_s)\right] \tag{52}$$

where $\delta \in (0,1)$ is the social discount factor, and $v(F_s)$ is the value of knowledge at time s.

When considering a specific research project, the social planner selects novelty d and success probability ρ to maximise:

$$\max_{d,\rho} \rho V_A(d; X_A) - \eta c(\rho, d) + \delta \mathbb{E}[W(F_{t+1})] \tag{53}$$

where $V_A(d; X_A)$ is the immediate benefit of discovering a question with novelty d given AI capability X_A, and $\mathbb{E}[W(F_{t+1})]$ is the expected continuation value.

The expected continuation value is:

$$\mathbb{E}[W(F_{t+1})] = \rho W(F_t \cup \{(x, y(x))\}) + (1-\rho) W(F_t) \tag{54}$$

Let $\Delta W = W(F_t \cup \{(x, y(x))\}) - W(F_t)$ represent the expected change in continuation value from a successful discovery. The social planner's first-order conditions are, with respect to ρ:

$$V_A(d; X_A) + \delta \Delta W = \eta c_\rho(\rho, d) \tag{FOC-SP-ρ}$$

and, with respect to d:

$$\rho \frac{\partial V_A(d; X_A)}{\partial d} + \delta \rho \frac{\partial \Delta W}{\partial d} = \eta c_d(\rho, d) \tag{FOC-SP-d}$$

In contrast, the scientist's first-order conditions are, with respect to ρ:

$$V_A(d; X_A) = \eta c_\rho(\rho, d) \tag{FOC-R-ρ}$$

and, with respect to d:

$$\rho \frac{\partial V_A(d; X_A)}{\partial d} = \eta c_d(\rho, d) \tag{FOC-R-d}$$

Comparing the social planner's and scientist's first-order conditions, we observe that the social planner's choices internalise two additional terms:

1. $\delta \Delta W > 0$ in the FOC for ρ, which increases the marginal benefit of a higher success probability;

2. $\delta \rho \frac{\partial \Delta W}{\partial d} > 0$ in the FOC for d, which increases the marginal benefit of novelty.

Following Carnehl and Schneider (2025), we know that without AI, these differences lead to the social planner choosing greater novelty and higher success probability than scientists: $d^s(\infty) > d^\eta(\infty)$ and $\rho^s(\infty) > \rho^\eta(\infty)$.

We now prove each part of the proposition:

Part (a): $X_A < X_E^*$ When $X_A < X_E^*$, we know from Proposition 4 that scientists choose the same novelty as in the no-AI case: $d^\eta(X_A) = d^\eta(\infty)$. This occurs because the benefits function is unchanged for $d > X_A$, and the scientist's optimum is achieved at $d^\eta(\infty) > X_A$. Similarly, for the social planner, the benefits function is unchanged for $d > X_A$, and the social optimum remains at $d^s(X_A) = d^s(\infty) > d^\eta(\infty)$. Therefore, the gap between the scientist's and social planner's choices remains the same as in the no-AI case.

Part (b): $X_A \in [X_E^*, d^s(\infty))$ When $X_A \geq X_E^*$, we know from Proposition 4 that scientists choose novelty $d^\eta(X_A) = X_A$. This occurs because the benefits function exhibits a discontinuity at $d = X_A$, making it optimal for scientists to choose exactly $d = X_A$.

For the social planner, there are two cases to consider:

- Case 1: The social planner might choose $d^s(X_A) = X_A$ if the discontinuity in benefits at $d = X_A$ is sufficiently large.
- Case 2: The social planner might maintain $d^s(X_A) = d^s(\infty)$ if the additional dynamic benefits of greater novelty outweigh the benefits of the discontinuity at $d = X_A$.

Given that $X_A < d^s(\infty)$ in this range, and that the social planner places greater value on novelty than scientists do, the social planner will maintain $d^s(X_A) = d^s(\infty)$ (Case 2). The gap between the scientist's and social planner's choices is now $d^s(\infty) - X_A < d^s(\infty) - d^\eta(\infty)$, which decreases as X_A increases.

Part (c): $X_A \geq d^s(\infty)$ When $X_A \geq d^s(\infty)$, the discontinuity in benefits at $d = X_A$ occurs beyond the social planner's preferred novelty without AI. In this case, both the scientist and the social planner choose $d = X_A$.

To prove this, note that for the scientist, the earlier argument from Part (b) still applies: $d^\eta(X_A) = X_A$. For the social planner, we need to compare the value of choosing $d = X_A$ versus $d = d^s(\infty)$. At $d = X_A$, the benefits are: $B(X_A) = \rho X_A + \delta \Delta W(X_A)$. At $d = d^s(\infty)$, the benefits are: $B(d^s(\infty)) = \rho V(d^s(\infty); \infty) + \delta \Delta W(d^s(\infty))$.

Since $X_A \geq d^s(\infty)$ and $V_A(d; X_A) = d$ for $d \leq X_A$, we have $X_A > V(d^s(\infty); \infty)$. The dynamic benefits ΔW are also at least as large for X_A as for $d^s(\infty)$ when $X_A \geq d^s(\infty)$, because a larger area of knowledge becomes accessible to AI, creating better future research opportunities.

Therefore, $V(X_A; X_A) > V(d^s(\infty); \infty)$, and the social planner chooses $d^s(X_A) = X_A = d^{(X_A)}$, eliminating the gap between private and social choices. This completes the proof of all parts of the proposition.

A.8 Proof of Proposition 7: Success Probability with AI

We analyse how AI capabilities affect the gap between the scientist's and social planner's choices of success probability.

A.8.1 First-Order Conditions

Recall that the social planner's FOC with respect to success probability ρ is:

$$V_A(d; X_A) + \delta \Delta W = \eta \, c_\rho(\rho, d) \quad \text{(FOC-SP-}\rho\text{)}, \tag{55}$$

where

$$\Delta W = W(F_t \cup \{(x, y(x))\}) - W(F_t)$$

measures the continuation value gain from a successful discovery, and $c_\rho(\rho, d)$ is the derivative of the cost function with respect to ρ. In contrast, the scientist's FOC is:

$$V_A(d; X_A) = \eta \, c_\rho(\rho, d) \quad \text{(FOC-R-}\rho\text{)}. \tag{56}$$

Since $\delta \Delta W > 0$, the planner's condition implies that a higher ρ is needed; that is, the social planner's optimal success probability ρ^s exceeds the scientist's ρ^η.

A.8.2 Proof of Part (a): $\rho^\eta(X_A) < \rho^s(X_A)$

For any given novelty d, the scientist chooses ρ^η so that

$$V_A(d; X_A) = \eta \, c_\rho(\rho^\eta, d), \tag{57}$$

while the social planner chooses ρ^s so that

$$V_A(d; X_A) + \delta \Delta W = \eta \, c_\rho(\rho^s, d). \tag{58}$$

Because $\delta \Delta W > 0$, we have

$$\eta \, c_\rho(\rho^s, d) > \eta \, c_\rho(\rho^\eta, d).$$

Since $c_\rho(\rho, d)$ is strictly increasing in ρ, it follows that $\rho^s > \rho^\eta$. Thus, for all X_A, the scientist's optimal success probability is lower than that of the social planner:

$\rho^{\eta}(X_A) < \rho^{s}(X_A)$.

A.8.3 Proof of Part (b): $\rho^{\eta}(X_A) > \rho^{\eta}(\infty)$ for $X_A > X_E^*$

Without AI, the scientist's optimal condition is:

$$V(d^{\eta}(\infty);\infty) = \eta\, c_{\rho}(\rho^{\eta}(\infty), d^{\eta}(\infty)). \tag{59}$$

With AI capability $X_A > X_E^*$, the scientist optimally sets the novelty to $d^{\eta}(X_A) = X_A$, so that

$$V_A(X_A; X_A) = \eta\, c_{\rho}(\rho^{\eta}(X_A), X_A). \tag{60}$$

By Lemma 1, for $d \leq X_A$ the benefit function is $V_A(d; X_A) = d$. Hence,

$$X_A = \eta\, c_{\rho}(\rho^{\eta}(X_A), X_A). \tag{61}$$

In the no-AI case, the benefit function is $V(d;\infty) = d - \frac{d^2}{6q}$ (for $d \leq 3q$) and since $d^{\eta}(\infty) \in (2q, 3q)$,

$$d^{\eta}(\infty) - \frac{(d^{\eta}(\infty))^2}{6q} < d^{\eta}(\infty).$$

For values of X_A just above X_E^* (with $X_E^* \approx 2q$) we then have

$$\eta\, c_{\rho}(\rho^{\eta}(X_A), X_A) > \eta\, c_{\rho}(\rho^{\eta}(\infty), d^{\eta}(\infty)).$$

Moreover, because the marginal cost $c_{\rho}(\rho, d)$ is higher for larger novelty, we conclude that

$$\rho^{\eta}(X_A) > \rho^{\eta}(\infty).$$

Thus, for $X_A > X_E^*$, the scientist's optimal success probability increases relative to the no-AI case.

A.8.4 Combining the Results: Gap Reduction but Not Elimination

Parts (a) and (b) together show that:

(i) $\rho^{\eta}(X_A) < \rho^{s}(X_A)$ for all X_A, and

(ii) $\rho^{\eta}(X_A) > \rho^{\eta}(\infty)$ for $X_A > X_E^*$.

Moreover, a similar argument establishes that the social planner's optimal success probability increases with AI (i.e., $\rho^s(X_A) > \rho^s(\infty)$ for $X_A > X_E^*$). However, the proportional increase in $\rho^\eta(X_A)$ is larger than that in $\rho^s(X_A)$—since the term $\delta\Delta W$ in the planner's condition lessens the relative impact of the immediate benefit increase. Consequently, while the gap $\rho^s(X_A) - \rho^\eta(X_A)$ remains positive for all X_A, it is reduced compared to the gap in the no-AI case: $\rho^s(\infty) - \rho^\eta(\infty)$. This completes the proof of the proposition.

A.9 Proof of Proposition 8: AI Research Cycles

We demonstrate that when $X_A > X_c^\eta$, AI tools naturally induce research cycles without external intervention, and that these cycles enhance welfare. We proceed in four steps, establishing each component of the research cycle and its welfare implications.

Step 1: Initial expansion by distance X_A. When $X_A > X_E^*$, scientists optimally expand knowledge by exactly distance X_A rather than distance $d^\eta(\infty)$, as established in Proposition 4. This occurs because the benefits of enabling AI's perfect decisions throughout the newly created interval of length X_A outweigh the benefits of any other expansion distance. This expansion creates a new area $[x_t, x_t + X_A]$ where x_t is the frontier point at time t.

Step 2: Subsequent scientists choose to deepen knowledge. When $X_A > X_c^\eta$, the newly created area of length X_A exceeds the threshold above which scientists prefer deepening knowledge to expanding it. To prove this formally, recall that at time $t+1$, scientists choose between:

- Expanding knowledge beyond the frontier point $x_t + X_A$, with expected utility $u_R^E(X_A) = \max_{d>0,\rho\in[0,1]} \rho V_E(d; X_A) - \eta\tilde{c}(\rho)d$
- Deepening knowledge within the interval $[x_t, x_t + X_A]$, with expected utility $u_R^D(X_A; X_A) = \max_{d\in[0,X_A/2],\rho\in[0,1]} \rho V_D(d; X_A, X_A) - \eta\tilde{c}(\rho)\sigma^2(d; X_A)$

By the definition of X_c^η, when $X_A > X_c^\eta$, we have $u_R^D(X_A; X_A) > u_R^E(X_A)$. Therefore, scientists at time $t+1$ strictly prefer deepening knowledge in the newly created area to continuing expansion.

Moreover, as shown in Proposition 5, when $X_A \in (X_c^\eta, 2X_A]$, scientists optimally position their discovery at $d^* = X_A/2$, which creates two equal subintervals of length $X_A/2$. This positioning is optimal because:

1. It minimises the variance of conjectures across the entire interval.
2. If $X_A \leq 2X_A$ (which is always true), both resulting subintervals $[x_t, x_t + X_A/2]$ and $[x_t + X_A/2, x_t + X_A]$ have length $\leq X_A$.
3. This enables AI to provide perfect decisions throughout the entire original interval.

This strategic positioning maximises the operational domain of AI, generating benefits of X_A (the entire interval length) rather than the smaller benefits that would accrue without strategic positioning.

Step 3: Return to knowledge expansion once knowledge is dense. After deepening at time $t+1$, knowledge consists of three points: $\{x_{t-1}, x_t, x_t + X_A/2, x_t + X_A\}$. This creates two intervals of length $X_A/2$, both of which are strictly shorter than X_c^η since $X_A/2 < X_A$ and $X_A > X_c^\eta$ implies $X_A/2 < X_c^\eta$.

By Proposition 4, when all available intervals have length less than X_c^η, scientists choose to expand knowledge beyond the current frontier. Therefore, at time $t+2$, scientists return to expansion, creating a new point at distance X_A from the current frontier: $x_{t+2} = x_t + X_A + X_A = x_t + 2X_A$.

This process then repeats: expansion by X_A, followed by deepening in the newly created area, followed by expansion again—creating the cyclical pattern described in the proposition.

Step 4: Welfare enhancement compared to ladder structure. We now demonstrate that this cyclical pattern enhances welfare compared to the ladder structure that emerges without AI or with $X_A < X_E^*$. The comparison involves two key components:

1. *Higher per-period value of knowledge:* After completing a cycle, knowledge is denser than in the ladder structure. Specifically, the ladder structure without AI creates intervals of length $d^\eta(\infty)$ with no deepening, while the AI-induced cycle creates intervals of length $X_A/2 < d^\eta(\infty)$. Since the value of knowledge

decreases with interval length beyond a certain point, this denser knowledge structure generates higher per-period value.

2. *Reduced failure risk:* The cyclical pattern strategically leverages both frontier and interior points to guide research. When deepening knowledge in the interval $[x_t, x_t + X_A]$, scientists utilise information from both boundary points to form more precise conjectures about interior points. This increases the success probability compared to frontier expansion (as established in Proposition 5), reducing the risk that the evolution of knowledge halts due to research failures.

To formalise the welfare comparison, let W_{cycle} and W_{ladder} denote the expected discounted welfare under the AI-induced cycle and ladder structures, respectively:

$$W_{\text{cycle}} = \sum_{j=0}^{\infty} \delta^{2j} \left[v(F_{2j}) + \delta \rho^D(X_A) v(F_{2j+1}) \right] \prod_{i=0}^{j-1} \rho^E(X_A) \rho^D(X_A) \tag{62}$$

$$W_{\text{ladder}} = \sum_{j=0}^{\infty} \delta^j v(F_j) \prod_{i=0}^{j-1} \rho^\eta(\infty) \tag{63}$$

where $\rho^D(X_A)$ and $\rho^E(X_A)$ are the optimal success probabilities for deepening and expanding with AI capability X_A, and $\rho^\eta(\infty)$ is the success probability without AI.

From Proposition 5, we know that $\rho^D(X_A) > \rho^E(X_A) > \rho^\eta(\infty)$ when $X_A > X_c^\eta$. Additionally, $v(F_{2j+1}) > v(F_{2j}) > v(F_j)$ due to the denser knowledge structure. Therefore, $W_{\text{cycle}} > W_{\text{ladder}}$, establishing that the AI-induced cyclical pattern enhances welfare.

This welfare enhancement arises endogenously from scientists' optimising behaviour in the presence of capable AI tools, without requiring external intervention—unlike the moonshot-induced cycles discussed in Carnehl and Schneider (2025) that require deliberate policy intervention.

A.10 Proof of Proposition 9: Optimal Funding Focus with AI

To establish the optimal focus for science funding with AI capability $X_A > X_E^*$, we analyse the marginal social returns to different policy interventions. We begin by examining the social planner's welfare maximization problem and identifying which

policy levers yield the highest welfare gains, taking into account both current and future periods through the social discount factor δ.

Recall that the social planner's welfare is: $W = \sum_{t=0}^{\infty} \delta^t v(F_t)$. The social planner considers how different policy interventions affect both the immediate value of knowledge and future knowledge accumulation. A policy intervention can influence scientists' choices of distance d and success probability ρ, as well as whether they expand or deepen knowledge.

Part (a): Prioritising funding for research at distance $d = X_A$. We first establish that when $X_A > X_E^*$, funding should prioritise research at distance $d = X_A$.

From Proposition 4, we know that scientists with cost parameter η choose $d^\eta(X_A) = X_A$ when $X_A > X_E^*$. From Proposition 6, we know that this choice of distance progressively aligns with the socially optimal novelty $d^s(\infty)$ as X_A increases, with perfect alignment when $X_A \geq d^s(\infty)$.

Consider a policy intervention that attempts to induce scientists to choose a distance $d \neq X_A$. We analyse two cases. The first case involves $d < X_A$. If scientists choose $d < X_A$, the immediate benefit is $V_E(d; X_A) = d < X_A = V_E(X_A; X_A)$. Thus, the immediate value of knowledge is strictly lower. Moreover, the knowledge structure created by choosing $d < X_A$ does not enable future scientists to make more valuable discoveries than would be possible with $d = X_A$. Therefore, both immediate and future welfare are reduced by encouraging $d < X_A$.

The second case is where $d > X_A$. If scientists choose $d > X_A$, the immediate benefit is:

$$V_E(d; X_A) = \begin{cases} d - \frac{d^2}{6q} & \text{if } X_A < d \leq 3q \\ [\text{decreasing in } d] & \text{if } d > 3q \end{cases} \tag{64}$$

For $d > X_A$, the AI cannot bridge the gap between existing knowledge and the new discovery, so the benefit reverts to the standard no-AI function. Since $V_E(d; X_A)$ is maximised at $d = 3q$ and decreases for $d > 3q$, there is no social benefit to encouraging $d > 3q$. For $d \in (X_A, 3q]$, the immediate benefit may be higher or lower than $V_E(X_A; X_A) = X_A$ depending on the specific values. However, scientists with cost parameter η choose $d = X_A$ precisely because the trade-off between benefits and costs favours this choice. Any policy intervention to induce $d > X_A$ would require substantial subsidies with diminishing welfare returns.

Moreover, choosing $d > X_A$ creates a knowledge structure where the newly cre-

ated area has length exceeding X_A, meaning the AI cannot provide perfect decisions throughout this area. This reduces the value of knowledge in future periods compared to creating an area of exactly length X_A where AI can operate perfectly. Therefore, funding should prioritize research at exactly distance $d = X_A$, reinforcing scientists' natural incentives in this regime rather than attempting to distort them.

Part (b): Subsidising increases in success probability ρ rather than further increases in novelty. Given that optimal distance is $d = X_A$, we now show that additional welfare gains come primarily from increasing the success probability ρ. From Proposition 7, we know that even when $X_A > X_E^*$, scientists choose $\rho^\eta(X_A) < \rho^s(X_A)$, where $\rho^s(X_A)$ is the socially optimal success probability. This gap between private and social success probabilities represents a remaining inefficiency in the knowledge creation process.

The marginal social benefit of increasing ρ is:

$$\frac{\partial W}{\partial \rho} = V_E(X_A; X_A) + \delta \Delta W \tag{65}$$

where ΔW represents the expected change in continuation value from a successful discovery. Since $\Delta W > 0$ (a successful discovery always improves future welfare), the social marginal benefit of increasing ρ exceeds the private marginal benefit $V_E(X_A; X_A)$.

In contrast, as established in part (a), there is a limited social benefit to increasing novelty beyond X_A. In fact, for $X_A \geq d^s(\infty)$, there is no social benefit whatsoever to increasing novelty beyond X_A.

Therefore, the marginal social return to subsidising increases in ρ exceeds the marginal social return to encourage further increases in novelty beyond X_A. Policy interventions should focus on addressing the underinvestment in research success probability rather than attempting to push novelty beyond the already optimal level.

Part (c): Encouraging deepening knowledge in areas of length $X \in (X_A, 2X_A]$ when $X_A < X_c^\eta$. Finally, we show that when $X_A < X_c^\eta$, there is an additional welfare benefit from encouraging deepening knowledge in areas of length $X \in (X_A, 2X_A]$.

From Proposition 5, we know that when $X_A < X_c^\eta$, scientists always choose to expand knowledge rather than deepen it, creating a uniform "stepping stone" pattern of knowledge evolution. While this pattern efficiently expands the frontier, it may

not optimally utilise AI's capabilities in existing areas.

Consider an area of length $X \in (X_A, 2X_A]$. Before intervention, scientists ignore this area and expand knowledge. However, as shown in Lemma 1 and Proposition 5, if a scientist deepens knowledge in this area by placing a discovery at position $d^* = X/2$, both resulting subintervals $[x_i, x_i + d^*]$ and $[x_i + d^*, x_{i+1}]$ have length $\leq X_A$. This means the AI can provide perfect decisions throughout the entire original interval after this strategic deepening.

The social benefit from such deepening is:

$$V_D(X/2; X, X_A) = X \tag{66}$$

This represents the maximum possible benefit (perfect decisions throughout an area of length X), which would not be achieved under scientists' natural incentives when $X_A < X_c^\eta$.

The key insight is that for areas of length $X \in (X_A, 2X_A]$, a single strategic discovery can transform a region where AI initially cannot operate (since $X > X_A$) into two regions where AI can operate perfectly. This creates extraordinary social value compared to the expanding knowledge pattern that would naturally emerge.

Therefore, when $X_A < X_c^\eta$, policy interventions should additionally encourage deepening knowledge in areas of length $X \in (X_A, 2X_A]$ to fully leverage AI's capabilities within existing knowledge domains.

Summary. The optimal policy focus with AI capability $X_A > X_E^*$ involves: (a) reinforcing scientists' natural incentives to choose distance $d = X_A$, (b) subsidizing increases in success probability to address the remaining gap between private and social optima, and (c) when $X_A < X_c^\eta$, additionally encouraging strategic deepening of knowledge in areas where doing so enables AI operation throughout previously inaccessible domains. This three-pronged approach maximises both immediate knowledge value and the long-term evolution of the knowledge frontier.

B Supplemental Appendix (Online): Phase Transitions in Research Strategy

This appendix provides a detailed analysis of the phase transitions that occur in research strategy as AI capability increases. These transitions represent qualitative shifts in how scientists optimally conduct research when AI tools of varying capabilities are available.

B.1 Proposition and Discussion

Proposition 10 (Phase Transitions in Research Strategy)**.** *As the AI capability parameter X_A increases:*

(a) At threshold $X_E^* \approx 2q$, a phase transition occurs in expanding strategy:

- *Below X_E^*: scientists expand knowledge ignoring the AI capability boundary*
- *Above X_E^*: scientists expand knowledge precisely to the AI capability boundary*

(b) This phase transition in expanding knowledge is characterized by:

- *A discontinuous jump in optimal distance from $d^\eta(\infty) < X_E^*$ to $d^* = X_A$*
- *A discontinuous increase in optimal success probability*
- *A fundamental shift in the pattern of knowledge evolution*

(c) For deepening knowledge, multiple phase transitions occur as the relationship between interval length X and AI capability X_A changes:

- *At $X = X_A$: AI transitions from operating across the entire interval to requiring deepening*
- *At $X = 2X_A$: The feasibility of enabling AI operation across the entire interval with a single discovery disappears*
- *At $X = 3X_A$ (when $X_A > \frac{3q}{4}$): Strategic AI-enabling placement becomes superior to standard deepening techniques*

These phase transitions represent qualitative shifts in research strategy that occur when AI capabilities reach critical thresholds. The transitions are discontinuous—small changes in X_A around these thresholds cause scientists to abruptly change both their distance choices and success probabilities. This creates bifurcations in the evolution of knowledge, with fundamentally different patterns emerging on either side of each threshold.

The most significant transition occurs at $X_E^* \approx 2q$ for expanding knowledge. Below this threshold, scientists operate essentially as they would without AI, choosing distances based on the traditional novelty-reliability trade-off. Above this threshold, scientists suddenly shift to expanding knowledge in perfect synchronisation with AI capability, choosing distances that precisely match the AI's operational range. This creates a distinctive stepping-stone pattern of knowledge evolution that would be unlikely to emerge without AI tools.

For deepening knowledge, the transitions are more complex and depend on the relationship between interval length and AI capability. When intervals are small relative to AI capability ($X \leq X_A$), no deepening is necessary. For intermediate intervals ($X_A < X \leq 2X_A$), strategic placement can enable AI operation throughout the entire original interval. For larger intervals ($X > 2X_A$), scientists must compare the benefits of enabling AI in parts of the interval against traditional deepening strategies.

These phase transitions have profound implications for science policy. Small improvements in AI capability around these thresholds can trigger large changes in scientific behaviour and knowledge accumulation patterns. This suggests that science funding and AI development efforts may have outsized impacts when they push AI capabilities across such critical thresholds.

In drug development, imagine AI systems initially having limited ability to interpolate between tested dosages (say, only within 1.5q). Scientists would ignore this limitation and continue choosing optimal dosages around 2.5q apart. However, if AI capabilities suddenly improved to interpolate across 2.5q or greater, scientists would dramatically shift their strategy, testing exactly at the AI's capability boundary and increasing their experimental precision.

Similarly, for deepening knowledge, consider a scenario where scientists know drug effects at doses of 0mg and 60mg. With an AI capability of $X_A = 25$mg, scientists couldn't enable AI operation across the entire interval with a single experiment. They

would strategically test at either 25mg or 35mg, enabling AI in at least one subinterval. If AI capability improved to $X_A = 30$mg, scientists would test at exactly 30mg, suddenly enabling AI operation throughout the original interval. These threshold effects create discontinuous reorganisations in how scientific knowledge accumulates.

B.2 Proof of Proposition 10

This section provides a formal proof of the phase transitions described above, analyzing both expanding and deepening knowledge scenarios and demonstrating the discontinuous changes that occur at specific thresholds.

Part (a): Threshold for Expanding Knowledge

We first establish the threshold X_E^* where scientists switch from Regime 2 (choosing $d^\eta(\infty)$ as in the no-AI case) to Regime 1 (choosing $d^* = X_A$ to precisely match the AI capability boundary).

The scientist's expected utility in Regime 1 (choosing $d \leq X_A$) is:

$$U_1^* = X_A[\rho_1^* - \eta\tilde{c}(\rho_1^*)] \tag{67}$$

where ρ_1^* satisfies the first-order condition:

$$1 = \eta\tilde{c}'(\rho_1^*) \tag{68}$$

The expected utility in Regime 2 (choosing $d > X_A$) is:

$$U_2^* = \rho_2^* \left[d^\eta(\infty) - \frac{(d^\eta(\infty))^2}{6q} \right] - \eta\tilde{c}(\rho_2^*) d^\eta(\infty) \tag{69}$$

where ρ_2^* satisfies:

$$1 - \frac{d^\eta(\infty)}{6q} = \eta\tilde{c}'(\rho_2^*) \tag{70}$$

and $d^\eta(\infty)$ satisfies:

$$\rho_2^* \left(1 - \frac{d^\eta(\infty)}{3q} \right) = \eta\tilde{c}(\rho_2^*) \tag{71}$$

The threshold X_E^* is defined as the value of X_A where these utilities are equal:

$$U_1^* = U_2^* \tag{72}$$

Solving for X_E^*:

$$X_E^* = \frac{U_2^*}{\rho_1^* - \eta\tilde{c}(\rho_1^*)} \tag{73}$$

Using the specific cost function $\tilde{c}(\rho) = (\text{erf}^{-1}(\rho))^2$, we can derive a more precise expression. Let $z_1 = \text{erf}^{-1}(\rho_1^*)$ and $z_2 = \text{erf}^{-1}(\rho_2^*)$. The first-order condition for ρ_1^* becomes:

$$1 = \eta\sqrt{\pi} z_1 \exp(z_1^2) \tag{74}$$

From this, we can express the denominator as:

$$\rho_1^* - \eta\tilde{c}(\rho_1^*) = \text{erf}(z_1) - \frac{z_1}{\sqrt{\pi}\exp(z_1^2)} \tag{75}$$

For typical parameter values where research costs are moderate (η is not too large), numerical analysis shows that $X_E^* \approx 2q$. This makes intuitive sense since:

1. The no-AI optimal distance $d^\eta(\infty) \in (2q, 3q)$.
2. When $X_A < 2q$, scientists find it optimal to explore beyond the AI capability boundary.
3. When $X_A > 2q$, scientists find it optimal to limit their exploration to precisely match the AI capability boundary.

Part (b): Characterisation of the Phase Transition

We now demonstrate the discontinuous nature of this phase transition.

Discontinuity in Distance. As X_A crosses the threshold X_E^*, the optimal distance jumps from $d^\eta(\infty)$ to exactly X_A. This represents a discontinuous change in research strategy, from ignoring the AI capability boundary to precisely targeting it.

Discontinuity in Success Probability. We show that $\rho_1^* > \rho_2^*$, indicating a discontinuous increase in research intensity at the threshold. From the first-order con-

ditions:

$$1 = \eta \tilde{c}'(\rho_1^*) \tag{76}$$

$$1 - \frac{d^\eta(\infty)}{6q} = \eta \tilde{c}'(\rho_2^*) \tag{77}$$

Since $d^\eta(\infty) > 0$, we have:

$$1 > 1 - \frac{d^\eta(\infty)}{6q} \tag{78}$$

Given that $\tilde{c}'$ is strictly increasing (due to the convexity of $\tilde{c}$), this implies $\rho_1^* > \rho_2^*$. Thus, scientists not only change their distance choice discontinuously but also increase their investment in research success probability when X_A exceeds X_E^*.

Shift in Knowledge Evolution. This transition fundamentally alters the pattern of knowledge evolution. Below X_E^*, knowledge accumulates in steps of $d^\eta(\infty)$, independent of AI capability. Above X_E^*, knowledge evolves in perfect synchronization with AI capability, accumulating in uniform steps of exactly X_A. This creates the distinctive stepping-stone pattern described in the text.

Part (c): Phase Transitions in Deepening Knowledge

For deepening knowledge within an interval $[x_0, x_1]$ of length $X = x_1 - x_0$, we identify several distinct phase transitions based on the relationship between X and X_A.

First Phase Transition: $X = X_A$. When X crosses X_A, the AI transitions from being able to operate across the entire interval to being unable to do so. This creates a discontinuity in the benefit function:

- For $X \leq X_A$: The AI operates across the entire interval without requiring deepening
- For $X > X_A$: Deepening becomes necessary to enable AI operation

Second Phase Transition: $X = 2X_A$. When X crosses $2X_A$, another discontinuity occurs:

- For $X_A < X \leq 2X_A$: Scientists can enable AI operation across the entire interval by choosing $d^* = X/2$

- For $X > 2X_A$: It becomes impossible to enable AI operation across the entire interval with a single discovery

The benefit function for deepening knowledge with AI is:

$$V_D(d; X, X_A) = \begin{cases} X & \text{if } d \leq X_A \text{ and } X - d \leq X_A \\ d & \text{if } d \leq X_A \text{ and } X - d > X_A \\ X - d & \text{if } d > X_A \text{ and } X - d \leq X_A \\ \frac{d(X-d)}{3q} & \text{if } d > X_A \text{ and } X - d > X_A \end{cases} \tag{79}$$

Optimal Deepening Strategy in Each Regime:

1. When $X \leq X_A$: Deepening offers no additional benefit beyond what the AI already provides.

2. When $X_A < X \leq 2X_A$: The optimal position is $d^* = X/2$ (assuming $X/2 \leq X_A$, which is true when $X \leq 2X_A$). This enables AI operation across the entire original interval, maximizing benefit.

3. When $X > 2X_A$: Scientists face three options:

   - Choose $d^* = X_A$, enabling AI in the left subinterval, with benefit $d^* = X_A$
   - Choose $d^* = X - X_A$, enabling AI in the right subinterval, with benefit $X - d^* = X_A$
   - Choose the no-AI optimal position $d^* = X/2$ (if $X \leq 6q$), with benefit $\frac{X^2}{12q}$

   Scientists compare these options and select the one with highest expected utility. This creates another phase transition at $X = 3X_A$, where the benefit from strategic AI-enabling placement (X_A) equals the benefit from standard deepening at the midpoint ($\frac{X^2}{12q}$). Solving $X_A = \frac{X^2}{12q}$ for X when $X = 3X_A$ yields $X_A = \frac{9X_A^2}{12q}$, which simplifies to $X_A = \frac{3q}{4}$. For $X_A > \frac{3q}{4}$, strategic AI-enabling placement is preferred for all $X \in (2X_A, 3X_A)$.

These multiple phase transitions reveal how the optimal deepening strategy changes discontinuously as the relationship between interval length X and AI capability X_A changes. Each transition represents a fundamental shift in how scientists approach knowledge deepening in the presence of AI tools.

References

Acemoglu, Daron, "Diversity and technological progress," in "The Rate and Direction of Inventive Activity Revisited," University of Chicago Press, 2011, pp. 319–356.

Aghion, Philippe, Benjamin F Jones, and Charles I Jones, "Artificial Intelligence and Economic Growth," in Ajay Agrawal, Joshua S Gans, and Avi Goldfarb, eds., *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, 2019.

Agrawal, Ajay, John McHale, and Alexander Oettl, "Finding needles in haystacks: Artificial intelligence and recombinant growth," in Ajay Agrawal, Joshua S Gans, and Avi Goldfarb, eds., *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, 2019.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?," in "Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency" ACM 2021, pp. 610–623.

Bryan, Kevin A and Jorge Lemus, "The direction of innovation," *Journal of Economic Theory*, 2017, *172*, 247–272.

Callander, Steven, "Searching and learning by trial and error," *American Economic Review*, 2011, *101* (6), 2277–2308.

Carnehl, Christoph and Johannes Schneider, "A Quest for Knowledge," *Econometrica*, 2025. forthcoming.

Castelvecchi, Davide, "Mathematicians welcome computer-assisted proof in 'grand unification' theory," *Nature*, 2021, *595*, 18–19.

Cockburn, Iain, Rebecca Henderson, and Scott Stern, "The Impact of Artificial Intelligence on Innovation," in Ajay Agrawal, Joshua S Gans, and Avi Goldfarb, eds., *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, 2019.

DeepMind, "AlphaFold: a solution to a 50-year-old grand challenge in biology," https://deepmind.com/research/case-studies/alphafold 2020. Accessed: 2025-02-27.

Doshi, Anil R and Oliver P Hauser, "Generative AI enhances individual creativity but reduces the collective diversity of novel content," *Science Advances*, 2024, *10* (28), eadn5290.

Gans, Joshua S, "Regulating the Direction of Innovation," Technical Report, National Bureau of Economic Research 2024.

_ , "Will user-contributed AI training data eat its own tail?," *Economics Letters*, 2024, *242*, 111868.

Hartnett, Kevin, "Proof Assistant Makes Jump to Big-League Math," 2021. Available online: https://www.quantamagazine.org/proof-assistant-makes-jump-to-big-league-math-20210610/.

Hopenhayn, Hugo and Francesco Squintani, "On the Direction of Innovation," *Journal of Political Economy*, 2021, *129* (7), 1991–2022.

Jumper, John, Richard Evans, Alexander Pritzel, and et al., "Highly accurate protein structure prediction with AlphaFold," *Nature*, 2021, *596* (7873), 583–589.

Kovalevskiy, Oleg, Robert A Nicholls, and Garib N Murshudov, "AlphaFold two years on: Validation and impact," *Proceedings of the National Academy of Sciences*, 2023, *120* (12), e2300376120.

Ludwig, Jens and Sendhil Mullainathan, "Machine learning as a tool for hypothesis generation," *Quarterly Journal of Economics*, 2024, *139* (2), 751–827.

Marcus, Gary, "The next decade in AI: four steps towards robust artificial intelligence," *arXiv preprint arXiv:2002.06177*, 2020.

Maynard, Alex, "Is AI Poised to Suck the Soul Out of Science?," Future of Being Human, Substack 2024. Accessed: 2025-02-27.

Mehta, Bhavik, "Formalising modern research mathematics in real time," 2023. Available online: https://xenaproject.wordpress.com/2023/05/03/formalising-modern-research-mathematics-in-real-time/.

Sloman, Leila, "'A-Team' of Mathematicians Proves a Critical Link Between Addition and Sets," 2023. Available online: https://www.quantamagazine.org/a-team-of-mathematicians-proves-a-critical-link-between-addition-and-sets-20231214/.

Song, Hyun Kyu, "Quo vadis experimental structural biology?," *Molecules and cells*, 2023, *46* (2), 71–73.

Toner-Rodgers, Aidan, "Artificial Intelligence, Scientific Discovery, and Product Innovation," 2024. Working paper, MIT.

Varadi, Mihaly, Stephen Anyango, Mandar Deshpande, and et al., "AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models," *Nucleic Acids Research*, 2022, *50* (D1), D439–D444.