

The Curse of Diligence? The Recognition Paradox of AI-Mediated Production

Wanhong HUANG
huangwanhong@serendip.ngo

Abstract

AI-mediated production changes the relationship between observable productivity and the evidence through which institutions recognize intellectual contributions. This paper investigates the conditional possibility that accelerated output may reduce substantive evaluative attention or weaken competence attributions even where underlying ability remains unchanged. The proposed explanation distinguishes output expansion, epistemic quality, attributional uncertainty, institutional recognition, and resource return. It examines evaluative congestion, possible recognition inversion, and the tragedy of productive diligence, in which acceleration, restraint, and additional verification can each impose costs. The project develops a bounded theoretical account, empirically discriminable hypotheses, a limited formal model, and candidate institutional responses. The empirical implications remain to be tested; no original causal estimates are reported.

Keywords: AI-mediated production; institutional recognition; productivity signaling; attributional uncertainty; trust; evaluative congestion

Discussion Paper Note

This discussion paper examines a possible mismatch between AI-mediated changes in knowledge production and the institutions that evaluate and recognize intellectual contributions. It treats productive capacity, artifact quality, competence, actual contribution, trust, recognition, and resource return as analytically distinct. The manuscript combines conceptual differentiation, a limited formal model, and empirically discriminable hypotheses. Its claims are restricted to the circumstances established by argument and available evidence.

The present source contains all 13 sections of a revised theoretical manuscript, with an unsectioned Introduction and Conclusion. The companion outline records its development and remaining scholarly review requirements. External references used in the text are recorded as checked against publisher or responsible-body records in the accompanying verification ledger. Illustrative cases are hypothetical, non-identifying, and institution-neutral. The theoretical propositions remain conditional and have not been empirically validated in this paper.

Responsible Use and Rights Reservation

This page provides a request for responsible scholarly use separately from the legal permissions of the licence. It does not amend the licence or statutory exceptions. The author welcomes scholarly critique, verification, teaching, and responsible adaptation. Users should preserve distinctions among hypothetical mechanisms, verified empirical findings, normative proposals, and demonstrated institutional effects. Claims concerning persons or institutions require evidence proportionate to their consequences.

The copyright notice, Creative Commons licence, and applicable permissions appear in the Notices that follow. This request for responsible use introduces no additional legal or technical restriction on uses permitted by that licence or by applicable law.

DRAFT

Notices

Publication status. Complete revised theoretical manuscript, subject to independent scholarly review and further language proofreading. This document is not peer-reviewed research and does not report validated original empirical results.

Licence. Except where otherwise indicated, copyright © 2026 Wanhong Huang. Original material is made available under Creative Commons Attribution–NonCommercial 4.0 International (CC BY-NC 4.0). The licence and applicable law govern permitted use. Full licence: <https://creativecommons.org/licenses/by-nc/4.0/>.

AI-assisted preparation. OpenAI’s ChatGPT assisted with conceptual organization, literature discovery, provisional drafting of Sections 1–13, six-language frontmatter translation, LaTeX adaptation, and editing. The human author determines the research object, theoretical commitments, interpretation, and any eventual publication, and remains responsible for accuracy and scholarly ethics. The six translations remain subject to independent review by proficient readers of the respective languages. ChatGPT is not identified as an author.

Research independence. The document reports no human-participant research and no original institution-specific findings. The use of the supplied source template does not imply endorsement by another project or institution.

Prefatory Address to Pioneers and Creators

Translations in the six official languages of the United Nations. The English text is the original version of this prefatory address.

English

We are living through a period that may prove critical in the evolution of knowledge and social life. The rapid development of artificial intelligence brings the limits of many inherited theories, including established formulations of social contract theory, into sharper view. Questions of subjectivity, the epistemology of generativity, the recognition of relational structures, and coexistence with vast heterogeneous relational systems arise alongside questions of trust, crises of symbolic meaning, global peace and development, energy, ecological stewardship and conservation, health and human development, the emergence of norms, and the understanding of justice.

Scholarship, too, has its generativity and its history. New forms of inquiry grow from earlier intellectual work, while changing conditions disclose questions that familiar concepts may be insufficient to address. A new era calls for pioneers and creators willing to recognize its emerging problems, develop new forms of knowledge, and undertake the patient work of testing and refining them.

I offer my deepest respect to all who devote their efforts to changing this world. Such work is rarely simple. I know something of the difficulties that some creators face.

With a well-equipped laboratory, capable assistants, and sufficient funding, a researcher may be able to devote sustained attention to a project, develop it coherently, and carry it through to completion. The circumstances are profoundly different for someone who must consider how to live for a day on two US dollars, while still striving to bring something into the world: something that shows they are still here, still alive, and still capable of creating and contributing. This, too, is a dilemma of pioneers and creators. The conditions needed to sustain creation can themselves become precarious in the course of creating.

I hope that diligence may receive the recognition, care, and positive return that allow it to continue. I wish every pioneer and creator, regardless of institutional position or the prominence of any title, the freedom to find happiness in their own way, and the kindness of a world that treats them well.

Français

Nous traversons une période qui pourrait constituer un moment critique dans l'évolution des savoirs et de la vie sociale. Le développement rapide de l'intelligence artificielle fait apparaître plus nettement les limites de nombreuses théories héritées du passé, notamment certaines formulations établies de la théorie du contrat social. Les questions de la subjectivité, de l'épistémologie de la générativité, de la reconnaissance des structures relationnelles et de la coexistence avec de vastes systèmes relationnels hétérogènes se posent aux côtés de celles de la confiance, des crises de la signification symbolique, de la paix et du développement dans le monde, de l'énergie, de la protection et de la conservation écologiques, de la santé et du développement humain, de l'émergence des normes et de la compréhension de la justice.

La recherche savante possède elle aussi une générativité et une histoire. De nouvelles formes de recherche naissent de travaux intellectuels antérieurs, tandis que l'évolution des conditions fait surgir des questions auxquelles les concepts familiers ne suffisent pas toujours à répondre. Une époque nouvelle appelle des pionnières, des pionniers et des créateurs capables de reconnaître ses problèmes émergents, de développer de nouvelles formes de savoir et d'entreprendre le travail patient de leur mise à l'épreuve et de leur perfectionnement.

J'adresse mon plus profond respect à toutes celles et à tous ceux qui consacrent leurs efforts à transformer ce monde. Une telle entreprise est rarement simple. Je connais quelque chose des difficultés auxquelles certains créateurs sont confrontés.

Avec un laboratoire bien équipé, des assistants compétents et des financements suffisants, une personne qui mène des recherches peut consacrer une attention soutenue à un projet, le développer avec

cohérence et le mener à son terme. La situation est profondément différente pour celle ou celui qui doit se demander comment vivre une journée avec deux dollars américains, tout en s'efforçant encore d'apporter quelque chose au monde : quelque chose qui témoigne de sa présence, du fait qu'elle ou il est encore en vie et demeure capable de créer et de contribuer. Tel est aussi l'un des dilemmes des pionniers et des créateurs. Les conditions nécessaires à la poursuite de la création peuvent elles-mêmes devenir précaires au cours de l'acte de créer.

J'espère que la persévérance recevra la reconnaissance, l'attention et les retours positifs qui lui permettront de se prolonger. Je souhaite à chaque pionnière, à chaque pionnier et à chaque créateur, indépendamment de sa position institutionnelle ou de l'importance de son titre, la liberté de trouver le bonheur à sa manière et la bienveillance d'un monde qui le traite avec égards.

Español

Vivimos un período que podría resultar decisivo en la evolución del conocimiento y de la vida social. El rápido desarrollo de la inteligencia artificial pone de manifiesto con mayor claridad los límites de muchas teorías heredadas, incluidas determinadas formulaciones consolidadas de la teoría del contrato social. Las cuestiones de la subjetividad, la epistemología de la generatividad, el reconocimiento de las estructuras relacionales y la convivencia con enormes sistemas relacionales heterogéneos surgen junto a las cuestiones de la confianza, las crisis del significado simbólico, la paz y el desarrollo mundiales, la energía, el cuidado y la conservación ecológicos, la salud y el desarrollo humano, la emergencia de normas y la comprensión de la justicia.

El conocimiento académico también posee generatividad e historia. Las nuevas formas de investigación nacen de trabajos intelectuales anteriores, mientras que las condiciones cambiantes plantean preguntas para las que los conceptos habituales pueden resultar insuficientes. Una nueva época necesita personas pioneras y creadoras dispuestas a reconocer sus problemas emergentes, desarrollar nuevas formas de conocimiento y emprender la labor paciente de ponerlas a prueba y perfeccionarlas.

Expreso mi más profundo respeto a todas las personas que dedican sus esfuerzos a transformar este mundo. Esa labor rara vez es sencilla. Conozco, al menos en parte, las dificultades que afrontan algunas personas creadoras.

Quien cuenta con un laboratorio bien equipado, asistentes competentes y financiación suficiente puede dedicar una atención sostenida a un proyecto de investigación, desarrollarlo de manera coherente y llevarlo a término. Muy distinta es la situación de quien debe pensar cómo vivir un día con dos dólares estadounidenses y, aun así, se esfuerza por aportar algo al mundo: algo que demuestre que sigue aquí, que sigue con vida y que todavía puede crear y contribuir. Este es también uno de los dilemas de quienes abren caminos y crean. Las propias condiciones necesarias para sostener la creación pueden volverse precarias durante el proceso creativo.

Espero que la dedicación reciba el reconocimiento, el cuidado y la respuesta positiva que permitan sostenerla. Deseo a todas las personas pioneras y creadoras, sea cual sea su posición institucional o la importancia de sus títulos, la libertad de encontrar la felicidad a su manera y la bondad de un mundo que las trate bien.

Русский

Мы живём в период, который может оказаться переломным для развития знания и общественной жизни. Стремительное развитие искусственного интеллекта всё отчётливее обнаруживает пределы многих унаследованных теорий, в том числе устоявшихся формулировок теории общественного договора. Вопросы субъектности, эпистемологии генеративности, познания реляционных структур и сосуществования с огромными гетерогенными реляционными системами встают наряду с вопросами доверия, кризисов символического значения, мира и развития во всём мире, энергетики, экологической ответственности и охраны природы, здоровья и человеческого развития, возникновения норм и осмысления справедливости.

Исследовательская деятельность также обладает генеративностью и собственной историей. Новые формы исследования вырастают из предшествующего интеллектуального труда, а меняющиеся условия выявляют вопросы, для решения которых привычных понятий может оказаться недостаточно. Новая эпоха нуждается в первопроходцах и созидателях, готовых распознавать возникающие проблемы, разрабатывать новые формы знания и терпеливо проверять и совершенствовать их.

Я выражаю глубочайшее уважение всем, кто посвящает свои усилия изменению этого мира. Такой труд редко бывает простым. Мне в некоторой мере знакомы трудности, с которыми сталкиваются некоторые созидатели.

При наличии хорошо оснащённой лаборатории, способных помощников и достаточного финансирования исследователь может надолго сосредоточиться на одном проекте, последовательно развивать его и довести до завершения. Совсем иная история — когда человеку приходится думать, как прожить день на два доллара США, и при этом продолжать стремиться привнести в мир что-то новое: нечто, свидетельствующее о том, что он по-прежнему здесь, жив и способен создавать и вносить свой вклад. В этом тоже состоит одна из дилемм первопроходцев и созидателей. Условия, необходимые для продолжения творческого труда, могут становиться всё менее устойчивыми в самом процессе созидания.

Я надеюсь, что усердие будет встречать признание, заботу и положительный отклик, которые позволят ему продолжаться. Желаю каждому первопроходцу и созидателю, независимо от институционального положения и значимости звания, свободы находить счастье по-своему и доброты мира, который будет относиться к нему достойно.

中文

我们正身处一个可能成为知识与社会生活演化关键转折点的时代。人工智能的迅速发展，使许多既有理论，包括社会契约论的一些传统表述，所面临的局限更加鲜明。主体性、生成性的认识论、关系结构的认识、与庞大的异构关系系统共生等问题，与信任、符号意义的危机、全球和平与发展、能源、生态关怀与保护、健康与人的发展、规范的涌现以及正义的认识等问题，一同摆在我们面前。

学问也具有生成性与历史性。新的探究形式生长于既有的思想劳动，而不断变化的条件又提出了熟悉概念未必足以回答的问题。新的时代需要先驱者与创造者，去认识正在出现的问题，发展新的知识形式，并耐心地检验、修正与完善它们。

谨向所有为改变这个世界而倾注努力的人致以最崇高的敬意。这样的工作很少轻而易举。我对一些创造者所面临的困境有所了解。

如果拥有设备完备的实验室、得力的助手和充足的经费，研究者或许能够持续专注于一个项目，有条理地发展它，直至完成。然而，如果一个人必须考虑如何用两美元度过一天，却仍努力为这个世界带来些什么——某种能够说明自己仍然在这里、仍然活着、仍然能够创造并作出贡献的东西——那便是完全不同的处境。这同样是先驱者与创造者的两难：支撑创造得以持续的条件，可能恰恰在创造的过程中变得岌岌可危。

愿每一份勤勉都能得到使其继续前行的认可、关怀与积极回应。也愿每一位先驱者与创造者，无论身处何种机构位置，无论头衔大小，都能够以自己的方式获得幸福，并受到这个世界温柔而善意的对待。

العربية

نعيش في مرحلة قد تكون منعطفًا حاسمًا في تطور المعرفة والحياة الاجتماعية. فالتقدم السريع للذكاء الاصطناعي يكشف بصورة أوضح حدود كثير من النظريات الموروثة، بما فيها بعض الصياغات الراسخة لنظرية العقد الاجتماعي. وتبرز أسئلة الذاتية، وإبستمولوجيا القدرة التوليدية، ومعرفة البنى العلائقية، والتعايش مع منظومات علائقية هائلة ومتغيرة، إلى جانب أسئلة الثقة، وأزمات المعنى الرمزي، والسلام والتنمية على الصعيد العالمي، والطاقة، ورعاية البيئة وصونها، والصحة والتنمية الإنسانية، ونشوء المعايير، وفهم العدالة.

وللعمل البحثي أيضًا قدرته التوليدية وتاريخه. تنشأ أشكال جديدة من البحث من الأعمال الفكرية السابقة، فيما تكشف الظروف المتغيرة عن أسئلة قد تعجز المفاهيم المألوفة عن معالجتها معالجة كافية. ويحتاج العصر الجديد إلى رواد ومبدعين مستعدين للتعرف إلى مشكلاته الناشئة، وتطوير أشكال جديدة من المعرفة، والاضطلاع بالعمل الصبور اللازم لاختبارها وتنقيحها وتحسينها.

أتقدم بأسمى عبارات الاحترام إلى جميع من يكرسون جهودهم لتغيير هذا العالم. فهذه المسيرة نادرًا ما تكون يسيرة. وأنا أعرف شيئًا من الصعوبات التي يواجهها بعض المبدعين.

قد يستطيع الباحث، حين تتوافر له مختبرات مجهزة جيدًا، ومساعدون أكفاء، وتمويل كافٍ، أن يكرس اهتمامه المستمر لمشروع واحد، وأن يطور به بصورة متماسكة حتى ينجزه. ولكن الوضع يختلف اختلافًا عميقًا بالنسبة إلى شخص يضطر إلى التفكير في كيفية تدبير يومه بدولارين أمريكيين، ومع ذلك يواصل السعي إلى تقديم شيء لهذا العالم: شيء يشهد على أنه ما زال هنا، وما زال حيًا، وما زال قادرًا على الإبداع والمساهمة. وهذه أيضًا إحدى معضلات الرواد والمبدعين؛ إذ قد تصبح الشروط اللازمة لاستمرار الإبداع نفسها هشة في أثناء عملية الإبداع.

أمل أن يقابل الاجتهاد بالتقدير والرعاية والاستجابة الإيجابية التي تتيح له الاستمرار. وأتمنى لكل رائد ومبدع، أيًا كان موقعه المؤسسي أو مكانة لقبه، الحرية في أن يجد سعادته بطريقته الخاصة، وأن يلقي معاملة طيبة من هذا العالم.

Contents

Discussion Paper Note	2
Responsible Use and Rights Reservation	3
Notices	4
Prefatory Address to Pioneers and Creators	5
1 Introduction	10
2 Problem Background and Analytical Definitions	12
3 AI-Mediated Production and Institutional Adjustment	18
4 Output Expansion and Evaluative Congestion	24
5 Attributional Uncertainty and Competence Inference	29
6 Recognition Inversion in AI-Mediated Production	35
7 The Tragedy of Productive Diligence	40
8 A Parsimonious Model of Production and Recognition	46
9 Recognition, Value Realization, and Resource Return	52
10 Institutional Adaptation and Evaluative Infrastructure	57
11 Empirical Research Designs and Competing Explanations	64
12 Discussion and Scope of Generalization	70
13 Conclusion	76
Research Ethics and Source Provenance	82

1 Introduction

AI-mediated production is changing the conditions under which intellectual work is performed, presented, and assessed. Writing assistants, code-generating systems, computational research agents, and other forms of machine assistance can alter the allocation of tasks among researchers, technical infrastructures, and evaluators. Their effects extend beyond the rate at which documents or software artifacts can be produced. They also concern the evidence through which an observer infers competence, the effort required to assess a contribution, and the institutional arrangements through which work acquires recognition and support. The present paper investigates a particular tension arising from these developments: an increase in a producer’s capacity to generate intellectual artifacts may coincide with diminished opportunities for substantive evaluation or with less favorable inferences about the producer’s competence. Under identifiable conditions, productivity that might ordinarily be taken as evidence of diligence can become a source of uncertainty or suspicion.

Evidence that AI assistance can alter productive performance is already available, although its scope varies markedly across tasks and organizational settings. In a randomized experiment on professional writing tasks, Noy and Zhang reported shorter completion times and higher assessed output quality among participants using ChatGPT [19]. In a field study of customer-support work, Brynjolfsson, Li, and Raymond found an average productivity improvement associated with access to a generative AI assistant, alongside considerable variation across workers and tasks [3]. These findings demonstrate that changes in the technical organization of work can affect realized performance. They do not establish equivalent effects for mathematical research, scientific validation, independent software engineering, or every form of creative inquiry. Nor does a measured increase in productivity by itself reveal how an external institution will assess the human contribution responsible for directing, checking, and applying the resulting work.

Scholarly production offers a particularly instructive setting because its outputs commonly pass through organized procedures of selection, review, publication, and attribution. A large-scale observational study spanning different forms of AI-augmented research, including methods beyond contemporary generative language models, reports associations with higher publication and citation measures for individual researchers, alongside narrower collective topic coverage and reduced subsequent engagement [12]. The reported associations challenge any general claim that AI-associated production necessarily attracts less recognition. At the same time, journal-level evidence indicates that expanded submission volume can place substantial pressure on evaluation. An analysis published by the *Organization Science* AI Task Force reports growth in submissions and changes in readability-based writing indicators and review characteristics in its observed corpus [8]. An editorial analysis in *Business & Information Systems Engineering* similarly distinguishes the falling costs of preparing submissions from the continuing costs of editorial screening, expert review, and certification [11]. These sources address different populations, outcomes, and evaluation stages. Together they motivate an examination of the conditions under which productive expansion, credibility assessment, and institutional reward diverge.

The recognition problem can be made precise by separating several objects that are often compressed into the single notion of productivity. The number of completed artifacts differs from their epistemic quality, practical usefulness, and potential contribution to a field. A producer’s underlying competence differs from the effort expended on a particular task and from the share of operations performed with computational assistance. The ability to explain and defend an artifact differs from possessing sole responsibility for every step in its fabrication. Finally, the attention an artifact receives, the confidence an evaluator assigns to its claims, the competence attributed to its producer, formal institutional recognition, and subsequent material return are distinct outcomes. They can vary independently. A paper may be judged technically sound without producing employment opportunities; a widely noticed software project may remain unverified; and a credible researcher may be unable to access an institutional channel capable of supporting further work.

The central recognition paradox concerns a possible reversal in the evidentiary meaning of productive diligence. Consider a hypothetical producer whose domain competence, intellectual judgment, and

responsibility for validation remain stable across two production arrangements. Under the second arrangement, AI assistance permits a larger quantity of outputs within a comparable interval. If evaluators continue to infer human competence from production-time assumptions developed for earlier workflows, an unusually large portfolio may be interpreted as weak evidence of individual ability, excessive reliance on automated generation, or an unacceptable burden of verification. Conversely, a smaller portfolio produced through familiar methods might appear more credible under the same evaluative conventions. The paradox arises when increased capacity and output are associated with less favorable judgments or reduced substantive scrutiny, even though the stipulated underlying competence has not declined. This is a conditional thought experiment rather than a report of an observed causal effect. Its empirical relevance depends on how evaluators respond to available quality evidence, task context, disclosure, institutional rules, and actual distributions of output quality.

The shift in evidentiary meaning can be approached through the economics of signaling, while allowing for differences between employment markets and scholarly evaluation. In signaling theory, observable characteristics can affect judgments when relevant capabilities are imperfectly observable [21]. The diagnostic value of a signal depends on its relationship to underlying attributes and on the process through which it is produced. AI changes some of those processes. A finished manuscript or functioning program may be easier to produce in certain respects, while generating a correct research question, identifying consequential assumptions, designing a valid test, and accepting responsibility for errors remain activities with distinct competence requirements. The apparent fluency or sheer quantity of artifacts may therefore become a less stable proxy for a particular human capability. This instability permits more than one response: evaluators may rationally request stronger evidence, rely excessively on coarse heuristics, adopt improved verification procedures, or transfer credibility from institutions and collaborators in the absence of direct examination.

Concerns about responsibility and verification have corresponding institutional expression. The International Committee of Medical Journal Editors recommends disclosure of AI-assisted technologies and locates accountability for accuracy and integrity with human authors [13]. These requirements reflect the legitimate need to know how claims were produced and who can answer for them. They also point to a distinction central to this inquiry: scrutiny warranted by a specific risk of error differs from an unsupported presumption that AI-assisted work is necessarily devoid of human competence or scientific value. An adequate evaluative system must be able to identify low-quality or misattributed work while remaining capable of recognizing rigorous contributions arising through changed productive mechanisms. Neither uncritical acceptance nor categorical depreciation provides a satisfactory basis for this task.

Evaluation capacity introduces a second mechanism. Production and evaluation draw on partially different resources, and their respective costs need not fall at the same rate. Even if the time required to prepare an artifact decreases, understanding its method, reproducing computations, identifying errors, and judging its significance can remain expensive. A rapidly growing population of submissions or demonstrations can consequently impose congestion on evaluators and compete for finite professional attention. Under some selection rules, the probability that any given work receives close examination may fall as the aggregate queue expands. This possibility should be distinguished from any proposition that the epistemic value of a valid result declines merely because similar artifacts become more numerous. The term *output inflation*, used provisionally in this paper, denotes a change in the volume and signaling environment of available outputs, with attention and evaluation pressures specified separately. It is an analytical analogy whose limits require explicit treatment.

A further difficulty arises when the producer attempts to respond to diminished attention or uncertainty through additional effort. Producing more artifacts can exacerbate portfolio congestion or attributional ambiguity. Reducing production can postpone potentially valuable inquiries, slow public dissemination, and forgo opportunities to develop reusable knowledge infrastructure. Documenting every generative step and arranging independent validation can improve assessability while consuming time, funding, and access that are unevenly distributed. Seeking institutional affiliation or endorsement can lower the costs faced by some evaluators, yet can also introduce dependence on established reputations and gatekeeping arrangements. The paper calls this configuration the *tragedy of productive diligence*: a

contingent predicament in which several reasonable responses to recognition difficulty carry costs that cannot be removed solely through greater individual exertion. The expression identifies a structure of competing constraints rather than a moral judgment about diligence or an assertion that all highly productive people experience disadvantage.

The recognition paradox also has implications beyond appraisal. Intellectual work depends on conditions that permit continuing participation, including access to time, tools, collaborators, employment, funding, and facilities for verification. Institutional acknowledgment can assist access to these resources, but it is neither sufficient nor always necessary. A validated contribution may remain economically unsupported because of limited demand, constrained budgets, eligibility criteria, intellectual-property arrangements, or the absence of a viable institutional interface for cooperation. Conversely, a producer may receive material support through channels imperfectly related to epistemic merit. Understanding productivity and recognition as separable processes is therefore a prerequisite for assessing whether additional production can sustain a creator's future work. The inquiry addresses the transition from produced artifacts to evaluative decisions and possible resource return while leaving a general theory of value circulation to subsequent research.

The argument proceeds through conceptual differentiation, institutional interpretation, a parsimonious formal representation, and proposals for empirical discrimination. It first establishes the problem background and definitions, then analyzes AI-mediated changes in productive mechanisms and relations. Subsequent sections examine evaluative congestion, contribution attribution, and conditional recognition inversion. A separate account of productive diligence identifies trade-offs across acceleration, restraint, disclosure, and institutional dependence. A formal model will specify assumptions under which output quantity, review selection, competence inference, and recognition can exhibit a nonmonotonic relationship. Proposed research designs will distinguish recognition effects from changes in actual quality, field-specific demand, selection bias, and justified evaluative caution. These analyses are presented as theoretical proposals to be assessed and revised, rather than as evidence of a universal recognition penalty.

Possible institutional responses are assessed with the same restraint. More legible records of generative processes can improve the inspectability of selected decisions and artifacts, provided that the burden of documentation remains proportionate to its evaluative purpose. Independent verification and opportunities for substantive discussion can provide stronger evidence than simple counts of outputs or disclosure labels. Evaluation institutions may also adapt their triage practices and invest in tools that lower the cost of checking claims while preserving responsibility for consequential judgments. Each response introduces its own distributional questions: who can afford verification, whose trajectory becomes visible, which authorities can confer recognition, and how evaluation translates into resources for continuing work. A comprehensive treatment of heterogeneous relational events, observer cognition, power configurations, and relational currency belongs to companion investigations; the present paper examines their interfaces with the particular recognition problem of AI-mediated production.

The contribution of this paper is thus a conditional account of how a transformation in productive capacity can alter the social interpretation of diligence without implying a corresponding deterioration in underlying competence or validated work. Its proposed recognition inversion remains an open empirical question. Its analytical importance lies in distinguishing genuine problems of quality and attribution from failures of attention, inference, and institutional adjustment. Framing these mechanisms separately permits a more precise investigation of when greater production supports recognition, when it fails to do so, and which evaluative arrangements may better sustain responsible participation in knowledge creation.

2 Problem Background and Analytical Definitions

The recognition paradox requires a precise account of the conditions under which intellectual production is assessed and the outcomes that evaluation can generate. The present section establishes the problem setting, introduces a hypothetical comparison that isolates a possible change in the interpretation

of productivity, and specifies the concepts used throughout the inquiry. The discussion distinguishes properties of outputs and producers from the evidence available to evaluators, and separates evaluative judgments from institutional decisions and material returns. Its purpose is conceptual and interpretive: the definitions organize later hypotheses without treating recognition inversion as an established empirical regularity.

2.1 Problem Setting and Historical Context

Intellectual production has long depended upon combinations of individual judgment, collective labor, specialized instruments, institutional resources, and methods of recording and transmitting results. The relative contributions of these components have varied across disciplines and historical settings. Printed texts, computational tools, laboratory instruments, collaborative software, and distributed research infrastructures have each altered both the feasible scale of production and the evidence available for assessing its authors. AI-mediated workflows introduce further changes, particularly where language generation, programming assistance, information retrieval, formal manipulation, and coordinated computational tasks become integrated into a producer's activities. The significance of these changes depends on the task, the reliability of the available tools, the competence of the producer, and the institutional practices governing the work.

Evidence for task-specific productivity effects provides an empirical starting point. Controlled professional-writing experiments and workplace studies report measurable changes in completion time and performance following access to generative assistance [19, 3]. The relevant outcomes and populations are specific to the studies concerned. Findings about routine writing or customer-support tasks do not establish the productivity, originality, or reliability of AI-assisted mathematical proofs, scientific experiments, or independent research software. Nevertheless, they establish that familiar relationships among time, labor input, and visible output can change with production technology. For an evaluator who lacks detailed information about the process, the evidentiary meaning of an artifact or portfolio may consequently become uncertain.

Institutional evaluation operates through distinct temporal and organizational constraints. Producing a draft, inspecting its claims, reproducing a computational analysis, assessing its originality, and determining its institutional significance are activities with different costs and participants. Scientific publication provides one setting in which those differences are visible. An editorial analysis of the publishing system distinguishes increasingly accessible manuscript production from the limited attention available for expert examination [11]. An account of a particular scholarly journal documents increased submission volume, declines in measured writing-readability indicators, and changes in review characteristics within its observed setting [8]. Such evidence warrants inquiry into review pressure; it does not imply that each additional AI-assisted artifact is low quality, nor does a case study of one journal establish conditions in every field.

The historical background matters because evaluative conventions commonly encode assumptions about how difficult particular outputs are to produce. In some settings, a long technical document may be regarded as evidence of sustained effort; in others, a functioning prototype may serve as a proxy for programming competence. Institutional affiliation, peer-reviewed publication, references from established collaborators, and prior grants may facilitate preliminary screening. These signals can convey genuinely useful information, while their diagnostic value depends on the circumstances of production, the integrity of the reporting process, and the relationship between the observed signal and the underlying attribute of interest. The general insight of signaling theory concerns inference from observable characteristics under incomplete information [21]; its application to contemporary research assessment remains an analytical extension requiring independent examination.

The institutional adjustment problem concerns the possible coexistence of rapidly changing productive arrangements and evaluative rules that adapt at different rates. Adjustment can occur through revised authorship standards, new documentation requirements, changes in reviewer practices, computational verification, or different definitions of contribution. Existing publication guidance, for example,

requires disclosure of relevant AI-assisted work while assigning responsibility for accuracy and integrity to human authors [13]. The presence of such changes makes a categorical thesis of unchanged institutions untenable. The relevant research question concerns the adequacy, accessibility, and distribution of these adjustments for different types of producers and outputs. Their consequences may differ between a resource-rich laboratory, an independent researcher, an applied engineering team, and a collective open-source project.

2.2 Presentation of the Recognition Paradox

A hypothetical comparison clarifies the paradox before its mechanisms are specified. Consider a researcher who develops computational methods in a technically demanding domain. During an initial observation interval, the researcher completes a modest portfolio through a workflow involving conventional programming and manual documentation. In a later interval of comparable duration, the researcher employs AI assistance for repetitive code generation, documentation, and preliminary exploration, thereby completing a substantially larger portfolio. The researcher continues to select problems, determine conceptual assumptions, test implementations, and accept responsibility for the results. For purposes of the comparison, assume that the researcher’s underlying competence is stable and that an independent retrospective assessment would find comparable epistemic quality among the outputs selected for comparison. The assumed quality comparison is a device for isolating attribution and evaluation; it does not assert that real AI-assisted portfolios invariably maintain quality as quantity increases.

An external evaluator may have access only to the artifact titles, abstracts, repositories, and reported dates of completion. The evaluator therefore faces two problems. First, examining the expanded portfolio at comparable depth requires additional time or a different allocation of review resources. Second, the observed pace and form of production may no longer support the competence inferences that the evaluator would have drawn from similar artifacts under earlier production arrangements. A large portfolio might prompt interest and substantive scrutiny. It might also be passed over because of review costs or discounted on the assumption that most of the intellectual work was automated. The same quantity signal can thus support different judgments depending on the evaluator’s prior beliefs, available evidence, task knowledge, institutional incentives, and capacity to verify the work.

The comparison isolates several possible outcomes.

The evaluative target must also be fixed. “Competence” can refer to independent performance without computational assistance, skill in directing and verifying an assisted workflow, capacity for original problem formulation, or sustained responsibility for a research program. Those constructs overlap in some circumstances and diverge in others. Evidence that strongly supports one may legitimately provide weak evidence for another. A recognition-inversion comparison is therefore interpretable only relative to a named competence construct and an assessment task that makes that construct relevant. The paper’s hypothetical fixed-competence cases hold the selected target construct constant, rather than presuming that every conception of expertise is invariant across changes in productive technology. In one case, reviewers examine representative artifacts and find them sound, leading to increased confidence in the researcher’s competence. In another, reviewers allocate only limited attention because the number of artifacts exceeds their available time, leaving competence judgments largely unchanged. In a third, the production rate itself produces an unfavorable inference about human authorship or understanding, despite the stipulated competence and ex post quality remaining constant. Only the third case provides a direct illustration of an adverse attribution effect; the second exemplifies evaluative congestion. Either can contribute to reduced opportunities for recognition relative to what might otherwise have occurred, although the causal paths differ.

The term *recognition paradox* names the general tension between expanded productive capability and the possibility of weaker external recognition. A more specific phenomenon, *recognition inversion*, concerns changes in a particular recognition outcome as output quantity increases under specified conditions. Definition 2.1 offers a provisional formulation, to be formalized in Section 8.

Definition 2.1 (Conditional recognition inversion). For a specified producer, evaluator, assessment context, and recognition outcome, a conditional recognition inversion occurs when a comparison between productive arrangements associates an increase in observable output with a less favorable recognition outcome, while the producer’s relevant underlying competence and the evaluated outputs’ relevant epistemic quality are held fixed or adequately controlled. The definition concerns a possible comparative relationship; establishing a causal effect requires further assumptions and evidence.

The recognition outcome must be named explicitly. Reduced willingness to read a portfolio differs from lower perceived technical competence; a delayed decision differs from an unfavorable quality judgment; and a failure to obtain funding differs from an absence of scholarly endorsement. Consequently, the paradox can arise through more than one mechanism and cannot be inferred from a single measure such as citations, publication counts, employment outcomes, or reviewer ratings. The thought experiment also leaves open the possibility that higher production results in more recognition. An adequate theory must accommodate both kinds of outcomes and identify the conditions that distinguish them.

2.3 Core Analytical Definitions

An intellectual output is a bounded communicable artifact or result produced within an identifiable workflow. In the present inquiry, examples include an academic manuscript, a formal derivation, an empirical analysis, a software module, a validated dataset transformation, or a documented technical design. The boundaries of an output are themselves conventional: one integrated research program may produce many records, while several computational contributions may be embedded in a single publication. *Output quantity* refers to the count or rate of artifacts under a declared counting rule and observation interval. Changes in packaging or publication practice can alter that count without changing the underlying substance. Comparisons of output growth therefore require explicit units and attention to segmentation.

Productive capacity denotes the range and rate of operations a producer and their accessible tools can perform under specified resource constraints. Capacity concerns feasible activity; realized output additionally depends on problem selection, organization, available inputs, verification costs, and external contingencies. *Productive diligence* denotes sustained purposive effort, care, and engagement in the relevant activity. This inquiry treats diligence as a descriptive and ethically significant category whose operational indicators are context dependent. It does not equate diligence with artifact counts, elapsed working hours, or the quantity of computational operations performed on a producer’s behalf. A high-volume portfolio can emerge through intensive human effort, extensive automation, teamwork, or mixtures of these conditions.

Epistemic quality concerns the properties that make an output suitable for responsible knowledge claims. Depending on the domain, these include logical validity, empirical adequacy, computational correctness, methodological transparency, robustness, originality, explanatory precision, and reproducibility. Quality is multidimensional, and a technically correct but unoriginal result may differ substantially from a novel but insufficiently supported claim. *Potential use value* identifies the capacity of an output to contribute to a relevant practical, scientific, educational, or collective purpose under specified conditions of access and use. Its realization may require further validation, compatible infrastructures, and willing adopters. Neither quality nor potential use value is directly represented by output quantity.

Competence refers to the task- and domain-relative abilities of an individual or collective to undertake specified intellectual operations, exercise judgment, diagnose errors, and respond to relevant challenges. A producer can possess competence without exercising every component of it in every artifact. *Actual contribution* refers to the identifiable intellectual and practical work undertaken by participants in producing an artifact, including problem framing, design, computation, validation, interpretation, and revision. *Attributable responsibility* concerns the obligations and authority associated with endorsing the artifact, correcting errors, and answering for claims. Contribution, competence, and responsibility may overlap without being interchangeable. AI assistance makes their relationships particularly important: the provenance of a generated passage reveals something about process, but alone establishes neither the absence nor the presence of the human capacities relevant to the work.

Table 1: Principal analytical objects in the production–recognition problem.

Object	Analytical focus	Indicative evidence or limitation
Output quantity	Count or rate under explicit artifact boundaries	Portfolio records; sensitive to packaging and time window
Productive competence	Domain- and task-specific capacity for judgment and performance	Explanation, problem solving, error diagnosis; imperfectly observed
Epistemic quality	Validity, adequacy, rigor, and related output properties	Independent checks, replication, domain-specific scrutiny
Contribution and responsibility	Work undertaken and accountability for claims	Process records, technical discussion, authorship declarations
Observability and legibility	Evidence accessible to and interpretable by an evaluator	Provenance, documentation, interaction, review costs
Recognition and resource return	Evaluation, credibility, institutional decisions, continuing support	Distinct outcomes with potentially different determinants

Observability denotes which portions of the production process, artifacts, validation records, and relevant histories can be accessed by a particular evaluator under given constraints. *Legibility* concerns the evaluator’s ability to organize accessible traces into an interpretable account of the relevant contribution. A detailed repository history may be observable yet difficult to interpret; a succinct experimental record may support a more reliable inference when its procedures and limitations are clear. *Evaluation capacity* refers to the resources available for selecting, inspecting, comparing, and validating claims within a stated interval and standard of scrutiny. Capacity is affected by expertise, time, infrastructures, workload, and institutional incentives. It varies with the complexity of the material and with the evidential threshold of the decision being made.

The analytic vocabulary distinguishes the production of an artifact from its examination and from the continued provision of resources to its producer. Table 1 summarizes the principal objects and the inferential distinctions necessary for later analysis. The table also identifies why simple output counts, surface fluency, or generalized reputation cannot substitute for the complete evaluative process.

The category of *output inflation* is used in a deliberately restricted sense. It refers to an increase in the supply of recognizable production artifacts relative to an explicitly specified benchmark, especially available evaluative capacity or the attention that a given artifact can expect to receive. The expression draws an analogy to the changing informational significance of abundant signals. It is not a claim that knowledge possesses a uniform price level or that the intrinsic correctness of a theorem diminishes when more theorems are published. Output inflation may coexist with major improvements in aggregate knowledge, with increasing low-quality material, or with both. The relevant distribution of quality must be investigated rather than inferred from the number of artifacts alone.

2.4 Trust, Recognition, and Resource Return

The term *trust* includes multiple relations that require separation. *Claim-directed epistemic confidence* concerns an evaluator’s degree of warranted confidence in a particular assertion or result. *Producer-directed credibility* concerns the weight an observer assigns to a source’s testimony, competence, or past reliability for a specified class of tasks. *Relational trust* concerns expectations about another party’s future conduct within an interaction, including reliability, responsiveness, and adherence to relevant commitments. These distinctions matter because a verified mathematical result can be accepted without general trust in its producer, while a reputable producer can issue a claim that warrants rejection. Confidence may also be calibrated differently across domains; expertise in one field is weak evidence of competence in an unrelated task.

Substantive evaluation refers to an attempt to assess specified properties of a contribution with methods proportionate to the decision, such as examination of assumptions, testing of executable claims, or comparison against relevant prior work. *Evaluative attention* denotes the allocation of scarce cognitive

and institutional resources to such assessment. A document can be encountered or cited without receiving substantive evaluation. *Institutional recognition* denotes a recorded or operative acknowledgment through procedures such as acceptance, certification, credentialing, selection, appointment, or adoption by an organization. The institutional standard may concern epistemic merit, practical fit, legitimacy, or a combination of criteria. Formal recognition and the underlying quality of a contribution are accordingly related through fallible procedures, rather than guaranteed to coincide.

Material resource return denotes resources that make subsequent participation possible, including income, paid positions, funded research time, equipment access, or committed organizational support. Return is specified relative to the actor and arrangement receiving it. It can arise through employment, contracts, grants, cooperative production, public provision, or other channels, each subject to distinct eligibility and distributional conditions. Academic recognition can contribute to such access, but an endorsed contribution need not yield income or continued funding. Conversely, remuneration can be granted for activities that are valuable to an organization without an independent scholarly endorsement of every artifact produced. The paper therefore treats recognition and material return as analytically separate dependent outcomes.

These outcomes interact over time. Successful independent validation can strengthen producer-directed credibility and encourage further examination of related work. Institutional recognition can improve the visibility of subsequent contributions. Resource access can expand production and verification capacity. Conversely, insufficient support can limit opportunities to assemble evidence even when the underlying competence is substantial. Such reciprocal processes motivate the later analysis of feedback and sustainability in Section 7. A complete account of the cognitive construction of trust, the distribution of power over recognition, and the circulation of relational credentials requires distinct theories. The present paper considers those processes where they bear directly on recognition under AI-mediated production.

The distinction between justified caution and a credibility discount requires particular care. An evaluator who identifies a false citation, unreproducible computation, or unsupported claim has evidence relevant to quality. Requests for accountable authorship and clear AI-use disclosure also have defensible purposes, as illustrated by established editorial guidance [13]. A different situation arises when AI involvement or high production volume alone is treated as sufficient evidence that the responsible human lacks understanding, irrespective of available tests and demonstrations. The latter inference may be inaccurate, although its actual frequency and consequences remain empirical questions. The theoretical object is the relation between the evidence consulted, the inference made, and the institutional outcome, with attention to reasonable alternatives and uncertainty.

2.5 Research Scope and Boundary Conditions

The primary domain of this investigation is AI-mediated knowledge production in settings where contributions can be inspected and potentially used by other persons or institutions. Scholarly manuscripts, computational research software, technical designs, and public research infrastructures are illustrative cases. The paper also draws cautiously upon task-specific empirical studies outside academic research to motivate production changes, with explicit limits on transferability. Its principal analytical unit is an evaluation relation involving a producer, an artifact or portfolio, an evaluator, and an institutional setting. The relevant human contribution may be individual or collective. The model does not require the producer to hold a conventional institutional appointment.

Several boundary conditions govern the proposed paradox. Increased productivity can increase both actual quality and recognized contributions, and additional AI-generated material may sometimes reduce quality. An increase in output can legitimately warrant different sampling rules or a longer evaluation process. Observed recognition differences may reflect changes in topic demand, evaluation standards, market conditions, collaboration, institutional eligibility, or portfolio composition. Recognition inversion should be identified only when these explanations are addressed adequately for the claim at issue. A change in competence attribution is analytically distinct from a change in opportunity resulting from

crowded review processes; the empirical designs developed in Section 11 will need to discriminate between them.

The investigation also preserves an institutional and normative distinction. A productive system may generate results of considerable public value while leaving its creators without stable support; that possibility motivates inquiry into resource-return interfaces without presuming a particular theory of distributive justice. Existing mechanisms of professional review serve necessary functions, and adaptation may require additional verification work as well as improved production tools. Normatively desirable outcomes must be evaluated against the protection of epistemic rigor, equitable access to assessment, accountable attribution, and the sustainable conditions of participation. No single metric of output or trust settles these questions.

The definitions in this section establish the objects of the subsequent mechanism analysis. Section 3 examines how AI alters productive mechanisms and their institutional interpretation. Sections 4–6 then investigate the pathways from greater output and attributional uncertainty to potentially divergent recognition outcomes. The formal and empirical components will identify their assumptions and possible falsifiers. At this stage, the recognition paradox remains a structured research question whose plausibility depends on the configuration of technical capacity, evidence, evaluative practices, and institutions.

3 AI-Mediated Production and Institutional Adjustment

This section identifies the productive changes that make the recognition problem possible. AI assistance can modify the feasible scale of work, the division of tasks, and the relationships through which responsibility and credit are assigned. The analysis proceeds from productive capacity to workflow organization and control, then examines the competence signals and evaluative conventions that institutions use to interpret those arrangements. The central issue is the pace and direction of adaptation: tools, labor arrangements, and assessment procedures need not change together.

3.1 Changes in Productive Capacity

Productive capacity denotes the feasible amount and range of work a producer can accomplish under specified constraints of time, knowledge, tools, infrastructure, and access to complementary labor. Within knowledge production, those constraints operate differently across activities. Drafting explanatory prose, generating executable code, developing an original mathematical argument, identifying an appropriate experimental design, and verifying a scientific claim impose distinct demands. Consequently, the productivity effects of an AI system should be specified at the level of particular tasks or connected workflows. An increase in the production of plausible artifacts establishes neither a corresponding increase in verified knowledge nor a uniform improvement in the underlying abilities of the person using the system.

Experimental evidence establishes that generative assistance can change the labor requirements of some activities. Noy and Zhang report improvements in task completion time and rated performance for professional writing tasks, while Brynjolfsson, Li, and Raymond report average productivity gains, with heterogeneous effects, in a field study of customer-support agents [19, 3]. These results provide evidence of technological changes to task performance under particular experimental and organizational conditions. Their applicability to independent scientific reasoning, research-software reliability, and disciplinary certification remains an additional research question. The relevant comparison is therefore between feasible sets of specified activities, rather than between universal quantities of intelligence or creativity.

An important consequence concerns the changing composition of production costs. A computational assistant may reduce the time required to construct an initial manuscript, propose a program, translate an algorithm into a second programming language, or explore alternative implementations. Costs associated with obtaining trustworthy inputs, specifying a consequential research question, checking mathematical assumptions, running experiments, identifying hidden dependencies, and interpreting results may remain

substantial. Some costs may even increase when rapid generation produces a larger set of candidates requiring examination. The production process then contains stages whose throughput expands at unequal rates. A workflow with inexpensive candidate generation and expensive validation can generate more provisional artifacts than it can responsibly certify. This asymmetry creates a reason to distinguish *generation throughput*, *verification throughput*, and *validated research output*.

The same distinction applies to the breadth of work available to an individual or small group. Access to computational assistance may reduce the cost of exploring problems across several fields, creating preliminary software interfaces, or organizing extensive bodies of technical literature. The review by Wang and colleagues documents the diverse roles of AI in scientific discovery, including work on hypotheses, experimental and computational processes, and scientific data, while discussing limitations associated with data and scientific understanding [25]. Such capabilities can enlarge the space of possible inquiries without eliminating field-specific prerequisites. A numerical simulation requires valid boundary conditions and an interpretable model; an analytical pipeline requires attention to measurement and data provenance; a scientific algorithm must satisfy tests appropriate to its intended use. Increased exploratory scope and greater verified competence are therefore analytically independent, even though they may support one another over time.

Productive-capacity changes can also be discontinuous in practice. Access to a new tool may make a previously infeasible workflow manageable, enable repeated revisions within a limited period, or remove a specialized implementation bottleneck. These effects depend on the producer's ability to formulate tasks, recognize failure, and integrate successive results. In other settings, model unreliability, costs of correcting generated errors, infrastructure restrictions, or the lack of necessary empirical data may eliminate the expected gains. A theory of recognition should admit both possibilities. It should also recognize that changes in output volume may reflect altered prices of particular production stages, choices about which artifacts to release, or changes in resources, rather than a change in a producer's underlying intellectual capacity.

This heterogeneity matters for institutional evaluation. A completed artifact conveys information about its quality only insofar as relevant properties can be inspected. It conveys information about the producer's abilities through a further inference, which depends on how the artifact could have been generated. When the relationship between production time, technical assistance, and observable output changes, previously familiar cues may become less diagnostic. That possibility is independent of whether the particular artifact is excellent, adequate, or defective. The mechanisms by which institutions respond to a growing supply of artifacts are examined separately in Section 4.

3.2 Transformations in Productive Mechanisms

The organization of AI-mediated work requires a more differentiated account than a contrast between autonomous human production and automated machine production. Intellectual projects combine a sequence of partially separable operations: selecting a problem, specifying relevant concepts, choosing a methodology, generating candidate arguments or artifacts, testing claims, interpreting findings, and revising the work in response to failure or criticism. Different participants and tools may contribute to different operations, and their contributions may vary across iterations. The product's final appearance does not, by itself, disclose how those operations were organized.

For analytical purposes, the production process can be represented as a set of functions whose allocation is variable across settings. A researcher may originate a conceptual hypothesis, ask a model to enumerate counterexamples, implement the associated computational test with model assistance, independently inspect the resulting code, and revise the initial hypothesis when the test fails. Another researcher may accept a generated explanation without sufficient examination. The two workflows can produce artifacts with comparable surface fluency while differing substantially in reasoning, error control, and accountable contribution. Equally, different workflows can produce equally valid final results despite unequal distributions of labor among humans and computational systems. Any general evaluation rule inferred from AI involvement alone would overlook these distinctions.

AI assistance can change the *sequence* of inquiry as well as the costs of its individual operations. A producer may move iteratively among theoretical exploration, implementation, critique, and redesign within a short interval. Rapid feedback can permit the construction of provisional representations that would previously have required additional personnel or more time. Yet iteration creates its own dependencies. Repeatedly asking the same model family to critique earlier generated material may yield correlated errors; successive code revisions can propagate unexamined assumptions; and apparently independent outputs may share training-derived patterns. The number of iterations is consequently an inadequate measure of epistemic independence. The appropriate evidentiary unit is the procedure through which consequential claims were checked and, where possible, subjected to independent constraints.

The distinction between reproduction and independent corroboration is useful here. The National Academies' consensus report distinguishes reproducibility, in the sense of obtaining consistent computational results using the same data and procedures, from replicability, which concerns consistent results across studies addressing the same question with new data or other independently obtained evidence [17]. The terminology has domain-specific applications, but the methodological lesson is general: a functioning script, a repeated model-generated explanation, and an independently corroborated substantive claim supply different kinds of evidence. A credible production mechanism should make relevant differences inspectable without pretending that every project needs the same verification method.

Task allocation also introduces a question of responsibility. The person presenting a manuscript may be responsible for its claims while relying on other parties or systems for particular computational operations. Responsibility for the claims and the factual distribution of work are different attributes. Some systems provide suggestions that are easy to test and reverse; others influence modeling assumptions, literature selection, or interpretative judgments in ways that are more difficult to detect. The International Committee of Medical Journal Editors accordingly locates publication accountability with human authors and requests disclosure of AI-assisted technology under its recommendations [13]. This example illustrates an institutional response to a changed division of epistemic labor, while its specific requirements should not be treated as binding rules for unrelated disciplines.

For the present paper, the central methodological point is that *human understanding*, *human contribution*, *technical assistance*, and *accountability* need separate evidence. A producer might possess the competence to assess a generated result without having manually executed each step; conversely, a producer might generate or submit an artifact whose claims they cannot adequately defend. Recognizing such distinctions helps evaluators maintain scrutiny without assuming either that extensive AI assistance negates human competence or that polished output demonstrates it. Section 5 will investigate the inferential problems associated with these distinctions in greater detail.

3.3 Transformations in Productive Relations

Productive relations concern the historically situated allocation of access, control, obligation, benefits, and decision-making authority among participants in a production process. In AI-mediated research, relevant participants may include researchers, technical collaborators, model providers, data custodians, owners of computational infrastructure, funders, employing institutions, downstream users, and evaluators. These roles may overlap within a single organization or be distributed across several organizations. The significance of AI for production relations therefore extends beyond the delegation of drafting or programming tasks: it concerns the changing practical conditions under which productive capacity can be exercised and recognized.

A first dimension is access to means of production. Model access can depend on subscriptions, computational budgets, network availability, data permissions, and technical skills. Some activities require substantial local computational resources or expensive experiments, while others can be conducted with inexpensive tools and public data. AI assistance may enable comparatively resource-constrained producers to undertake work previously requiring larger teams. It may also create new dependencies on external services whose prices, usage limits, model capabilities, or conditions of access can change. The direction of the resulting distributive effect cannot be inferred from the presence of AI alone; it requires

examination of actual access conditions and the substitutability of complementary resources.

A second dimension is control over the production process. A researcher may decide what to study and when to release a result while having limited control over a proprietary model’s architecture, training data, or reproducibility across software versions. A collaborating institution may control access to research materials, experimental facilities, publication authorization, or intellectual-property decisions. Conversely, openly specified algorithms, locally executable tools, portable data, and documented interfaces may increase some kinds of autonomy. These arrangements can shape which contributions become technically feasible, independently checkable, and publicly accessible. They also complicate the identification of the individual actor to whom visible productivity should be attributed.

A third dimension concerns how benefits and burdens are distributed. AI-assisted production can generate value for the individual researcher, an employer, the model provider, the institution that controls downstream adoption, and users who build on the output. Those benefits need not return proportionately to the actor responsible for originating a question, evaluating candidates, or maintaining a public research artifact. The labor of verification may be assigned to authors, reviewers, software maintainers, or users, while decisions over funding and certification remain elsewhere. These distinctions motivate the later separation of recognition from material return in Section 9. A comprehensive theory of circulation across heterogeneous relational arrangements, however, remains outside this paper’s scope.

A fourth dimension is the allocation of epistemic authority. A funding body can decide which results warrant support; a journal can grant publication standing; a hiring committee can interpret evidence of competence; a repository maintainer can decide whether an external contribution meets its requirements. Such authority can be associated with relevant technical expertise, legitimate procedural responsibility, and genuine capacity to verify a claim. It can also create dependencies for producers who lack institutional standing. The relation between authority and trust is thus a question of how authority is constituted and exercised, as well as whether relevant claims can be inspected and contested. The detailed formation of power structures in evolving relational trajectories is reserved for a companion investigation; here the immediate concern is the capacity of evaluative institutions to recognize work arising from new production arrangements.

These dimensions interact. Expanded productive capacity may coexist with dependence on model infrastructure; greater independence in proposing research can coexist with dependence on established institutions for recognition; and open dissemination may increase the accessibility of outputs without creating a mechanism for financing their maintenance. A historically informed analysis must therefore keep technological capability, effective access, authority over evaluation, and material security analytically distinct. The theoretical recognition paradox is located within this wider organization of production, but it concerns a narrower question: how altered production arrangements affect judgments about a producer’s competence and contribution.

3.4 Evaluative Conventions and Competence Signals

Evaluative institutions operate under limited observation and must decide how much evidence is adequate for particular judgments. A manuscript, portfolio, application, demonstration, or repository can make only some aspects of its production visible. When examining a candidate, evaluators may rely on indicators such as the coherence of an argument, technical validation, previous publications, scholarly references, team membership, documented responsibilities, or sustained participation in a project. These indicators carry different information and entail different examination costs. An evaluative convention is a relatively stable practice through which an institution uses such indicators to infer attributes that cannot be observed directly.

One historically contingent convention links the volume and complexity of finished work to an expectation of time-consuming human labor. It can provide useful information in settings where production methods, team structures, and quality controls are broadly comparable. Yet the informational meaning of this convention can change when production costs or the distribution of work change. Under AI-assisted workflows, an extensive manuscript may require relatively little drafting time while still demanding so-

Table 2: Selected interfaces between AI-mediated production arrangements and institutional evaluation.

Production change	Evaluative question	Relevant evidence
Faster artifact generation	What does volume reveal about competence?	Independent explanation and task-specific performance
Model-assisted drafting or coding	Which steps were performed or verified by whom?	Contribution description, tests, and review records
Rapid iterative exploration	Are apparently distinct results evidentially independent?	Methods, data, provenance, and independent checks
External model and infrastructure reliance	Which dependencies constrain reproduction and responsibility?	Access terms, versions, execution conditions
Expanded cross-domain work	Which claims meet disciplinary validity requirements?	Domain-appropriate validation and expert scrutiny
Institutionally mediated endorsement	Does the signal reflect verification or inherited status?	Scope of examination and documented basis of endorsement

phisticated conceptual design and prolonged checking. Another extensive manuscript may be generated with little substantive oversight. Output size alone may consequently become a weak discriminator between these cases. The general signaling problem, originally developed in the context of labor-market uncertainty, offers a relevant analogy: an observable signal’s evidentiary significance depends on the process connecting it with the unobserved attribute [21].

Institutions also employ distinct concepts of performance. For research software, a successful build or demonstration indicates that some computational procedure executes under specified conditions; it provides less direct evidence of algorithmic correctness, scientific validity, maintainability, or the capacity to explain a surprising result. For a theoretical argument, a coherent presentation is distinct from a valid proof, a defensible interpretation, or a novel contribution to existing research. For collaborative work, an author’s position or institutional affiliation is distinct from the author’s actual intellectual responsibility. Treating each signal according to its domain of validity is an ordinary requirement of sound evaluation, irrespective of AI. The altered production environment makes this requirement more salient because text fluency, executable code, and polished presentation can be produced through a wider range of processes.

Evaluation therefore has two related but separable objects. *Artifact assessment* concerns the claims and properties of the submitted work: its correctness, evidential grounding, relevance, and usefulness for a specified purpose. *Producer assessment* concerns the abilities and dispositions of a person or team: problem formulation, understanding, judgment, reliability, and capacity to carry out future work. A highly reliable artifact can be useful regardless of the producer’s institutional status, while a hiring decision may reasonably require evidence of the candidate’s ability to respond to novel challenges. Confusion arises when evidence appropriate for one object is used as if it established the other. A successful AI-assisted artifact should not be rejected solely because it required less labor than anticipated; neither should its existence be treated as sufficient proof of independent human competence.

This distinction offers a basis for interpreting apparently contradictory responses to high productivity. Substantial output can increase an evaluator’s confidence when accompanied by reproducible tests, clear conceptual explanations, and credible records of responsibility. It can decrease confidence when there is direct evidence of fabrication, unreconciled contradictions, or failures of understanding. It can also generate unfavorable judgments under conditions where actual quality and competence remain unchanged but the evaluator interprets unusual output volume through an inaccurate assumption about its generation. Only the last configuration constitutes the specific recognition inversion studied here; the first two represent important positive and justified-negative comparison cases. Which configuration occurs is an empirical matter requiring observation of evaluators’ information, tasks, and decisions.

Table 2 summarizes the relation between selected production changes and the evaluative questions they make salient. The entries identify analytical interfaces rather than established causal outcomes.

3.5 Uneven Institutional Adaptation

The relation between changing production methods and changing evaluation practices is best understood as heterogeneous institutional adjustment. Disciplines and organizations differ in the kinds of claims they assess, the risks associated with error, the availability of standardized tests, and the resources they can devote to evaluation. In some contexts, a computational artifact admits extensive automated checking; in others, the object of assessment is an interpretative argument, a context-dependent decision, or a long-term pattern of reliable conduct. The changes required to assess AI-mediated production responsibly are consequently neither uniform nor reducible to a single speed of institutional reform.

There is evidence that some institutions have already articulated differentiated practices. The International Committee of Medical Journal Editors assigns human accountability for AI-assisted scholarly content and requests disclosure under its authorship recommendations [13]. Springer Nature's editor guidance adopts a context-sensitive risk assessment that distinguishes comparatively low-risk assistance from uses affecting analysis and evaluative judgment, and from uses that compromise human accountability or research integrity [22]. These policies are institution-specific examples of adjustment. Their existence does not establish that they are implemented consistently, that all fields share their definitions, or that adherence invariably yields fair recognition. They do demonstrate that the institutional landscape includes active adaptation alongside continued uncertainty.

A useful comparison is the development of methodological standards for AI systems more generally. The National Institute of Standards and Technology's voluntary AI Risk Management Framework articulates context-dependent considerations of validity, reliability, accountability, transparency, and other attributes of trustworthy systems [23]. The framework is not a direct standard for judging scholarly authorship or producer competence. Its relevance is methodological: technical properties and social responsibilities require different but coordinated forms of examination. In scholarly settings, corresponding distinctions may involve checking the result, documenting consequential tool use, establishing who can explain a decision, and deciding which additional evidence is justified by the stakes of the claim.

Two mechanisms of delayed adjustment deserve separate consideration. *Procedural inertia* concerns the persistence of established screening rules, documentation requirements, and organizational workflows even when the evidentiary significance of their inputs changes. These practices may persist for legitimate reasons, including limited administrative capacity, the need for predictable treatment, and the protection of review integrity. *Epistemic uncertainty* concerns the genuine difficulty of determining which new methods produce reliable results and which indicators validly distinguish competence from automated presentation. In this case, cautious review can reflect a lack of adequate validated standards rather than resistance to technological change. The two mechanisms can coexist, but their remedies differ: revised procedures address the former, while better evidence, benchmarks, and validated assessment methods address the latter.

Institutional adjustment also redistributes evaluative burdens. A requirement to disclose consequential AI use can improve the information available to reviewers, yet an excessively detailed documentation demand may impose costs disproportionately on independent producers and small teams. Automated screening can help manage large volumes but may replicate biases or classify atypical contributions inadequately. Requiring a demonstration of personal understanding can provide relevant evidence in a hiring context, while becoming an intrusive or impractical demand if indiscriminately imposed on all publicly disseminated research. Any institutional remedy must therefore be evaluated jointly in terms of evidentiary benefit, administrative cost, proportionality, and distributive consequences. These questions will be developed in Section 10.

Countercases are essential to the argument. Some producers may gain justified recognition because AI-assisted work allows them to provide clearer explanations, more complete tests, or sustained public maintenance of research software. Some evaluators may already use artifact-specific verification that is largely insensitive to output quantity. Some domains may experience changes in production and validation capacity together, reducing rather than increasing evaluation pressure. Empirical research, using a broad classification of AI-augmented scientific activity that extends beyond generative language models, also reports associations with greater individual scientific impact alongside a narrowing of collective re-

search attention [12]. The study’s observational design and classification choices limit causal inference, but the coexistence of potentially favorable and unfavorable outcomes cautions against treating the recognition paradox as a universal description of AI-mediated research.

AI-mediated tools can therefore change what producers can accomplish, how their contributions are organized, and what an observer can infer from completed work. The next analytical question concerns the first institutional gate: whether a larger supply of artifacts can receive examination of comparable depth. Section 4 addresses that gate without treating increased output as a proxy for diminished knowledge quality.

4 Output Expansion and Evaluative Congestion

Output expansion affects recognition partly through the allocation of scarce review resources. This section distinguishes generated artifacts from validated contributions, represents the work required for substantive examination, and considers portfolio screening when review capacity grows more slowly than incoming demands. The expression *output inflation* is used as a limited analogy for changes in a signal’s discriminative value, with attention to its differences from monetary inflation. The mechanism studied here concerns selection into examination; the inference made *after* examination requires the separate attribution analysis of Section 5.

4.1 Throughput and Production-Cost Heterogeneity

The expansion of output begins with changes in what can be generated under a finite budget of time, labor, and computational resources. Yet a knowledge-production process consists of activities with different costs and epistemic roles. Producing a coherent draft, constructing a plausible hypothesis, collecting original observations, testing a proposed model, checking a proof, and maintaining a reproducible software system are distinguishable operations. Improvements in one operation can increase the number of candidate artifacts without increasing the number of validated contributions by the same proportion. A measure of output expansion therefore requires a specification of what counts as an artifact, when it enters an evaluative system, and which production operations its reported status includes.

Research on generative assistance provides evidence of productivity effects within particular task environments. In a controlled study of professional writing tasks, Noy and Zhang report gains in speed and assessed output quality; Brynjolfsson, Li, and Raymond find average productivity improvements in an organizational field setting, with heterogeneous effects across workers [19, 3]. These findings justify treating faster generation as a plausible feature of some AI-mediated workflows. Their tasks and measured outcomes differ from the verification of scientific claims. They do not determine the rates at which valid theorems, reproducible experimental findings, or sustainable software systems can be produced in other settings. The literature on AI-enabled scientific discovery likewise documents substantial differences among the activities to which AI is applied [25].

Costs also depend on the kind of artifact submitted for consideration. A short but consequential mathematical claim may require specialist verification, while a large body of routine documentation may be evaluated through a combination of automated checks and sampling. A functioning software demonstration can reduce the effort necessary to establish that a particular program executes, while leaving the scientific adequacy of its assumptions unresolved. The same number of artifacts can therefore create very different aggregate demands on evaluators. Counting submissions alone measures an important aspect of scale but provides an incomplete measure of the work required for responsible assessment.

For analytical clarity, let $\mathcal{A}_t$ denote the set of artifacts newly entering a particular evaluation environment during period t , and let w_i be the estimated amount of evaluation work required to reach a specified examination standard for artifact i . The aggregate incoming evaluation workload is defined in Equation 1.

$$W_t = \sum_{i \in \mathcal{A}_t} w_i, \quad w_i \geq 0. \quad (1)$$

The term w_i refers to evaluator effort under an explicitly stated standard, rather than the value, length, or correctness of an artifact. In practice, its estimation may itself be costly and uncertain. Equation 1 is accordingly an accounting representation useful for identifying evaluation demands; it is not a measured workload model or a claim that all assessments can be reduced to one uniform unit. Evaluation settings can require multiple dimensions of expertise, and the availability of one type of expertise may not compensate for a shortage of another.

It is additionally necessary to distinguish generated, submitted, examined, validated, adopted, and maintained artifacts. An author may generate many candidates but submit only a few; an institution may screen many submissions but examine few in depth; a valid result may receive little attention; a widely adopted result may require substantial continuing maintenance. The movement between these states depends on selection rules, evidence, and the resources available to participants. Increased generation throughput establishes a change in one part of this process. The recognition paradox concerns possible changes further along the chain, where other actors determine whether and how much of a producer's work they will examine.

4.2 Evaluation Capacity and Congestion

Evaluation capacity is the ability of a particular system to undertake assessments of specified scope and quality over a defined period. It includes available expert time, administrative coordination, access to relevant evidence, computational testing resources, and procedures for resolving contested assessments. Capacity cannot be inferred from reviewer headcount alone. Two institutions with equal numbers of evaluators may differ in technical specialization, triage design, infrastructure, incentives, and obligations to provide detailed feedback. They may also face different mixtures of artifacts. A fixed nominal capacity therefore offers only a provisional basis for comparison.

Let $C_t > 0$ represent available evaluation effort during period t , measured on the same conditional workload scale as W_t . Equation 2 defines an indicative incoming pressure ratio.

$$\rho_t = \frac{W_t}{C_t}. \quad (2)$$

A value of $\rho_t > 1$ indicates that the incoming workload exceeds the period's available effort under the specified examination standard. It does not establish a stable equilibrium, a particular waiting time, or a corresponding decrease in scientific quality. Institutions can reject submissions before full assessment, redirect specialized work, recruit additional reviewers, improve screening, reduce the depth of examination, or defer cases to subsequent periods. Each response has distinct implications for accuracy, fairness, waiting time, and the producer's opportunity to receive an explanation of the outcome.

The difference between resource constraints and actual congestion becomes clearer when previously unexamined work is considered. Let B_t denote the remaining workload at the beginning of period t and S_t the amount of evaluation work actually completed during that period, with $0 \leq S_t \leq C_t$. The simple bookkeeping relation in Equation 3 describes how pending workload can evolve.

$$B_{t+1} = \max\{0, B_t + W_t - S_t\}. \quad (3)$$

This relation assumes that submissions and uncompleted tasks can be represented as transferable units of pending evaluation work and that withdrawal or expiry is handled through the relevant accounting rules. It omits priority scheduling, specialist queues, variation in review quality, and strategic responses to growing delay. Its purpose is to make the capacity constraint explicit. Where the expected incoming workload persistently exceeds effective completion capacity, some combination of growing backlog, selective rejection, reduced examination depth, or institutional redesign becomes necessary.

Emerging research on scholarly publishing makes the workload distinction relevant. Gartenberg and colleagues report increased submission pressure and declines in writing-readability proxies and some review characteristics in a particular journal environment following the spread of generative AI; Hanneke and colleagues discuss how changing production costs can intensify demands on editorial screening and

review [8, 11]. These sources concern defined institutional settings and forms of scholarly publication. They neither measure the total stock of valuable knowledge nor establish a general causal effect of AI usage on the trustworthiness of individual contributors. Their value for the present argument lies in illustrating the possible divergence between artifact-generation capacity and institutional examination capacity.

A constrained evaluation system may ration scrutiny through methods that affect different producers unequally. It may favor a concise portfolio over a comprehensive one, prioritize submissions with recognizable institutional signals, use automated checks, allocate reviewer time according to expected disciplinary relevance, or request supplementary evidence from unconventional contributors. Some practices improve the allocation of scarce expertise; others may introduce distortions when their indicators cease to predict the attribute of interest. The resulting differences are properties of selection and assessment procedures. They should be studied separately from the intrinsic quality of submissions and from the motives attributed to individual evaluators.

4.3 Output Inflation as a Scoped Analogy

The expression *output inflation* names a possible decline in the discriminative value of readily produced output signals when their supply expands more quickly than opportunities for substantive evaluation. Its use is deliberately narrower than the monetary concept of inflation. Knowledge artifacts are heterogeneous, may be reused without depletion, and can generate additional value through replication or recombination. There is no general price index that converts their epistemic quality into a single commensurable magnitude. More publications or software artifacts cannot therefore be assumed to diminish the validity of existing work merely by increasing their number.

The analogy is useful at the interface between producers and observers. If a polished manuscript was historically costly to prepare, its appearance could sometimes function as weak evidence that the producer invested substantial time or possessed certain communicative abilities. When the cost of producing a polished manuscript changes, the same appearance can remain useful for reading while becoming a less informative signal of the producer's independent reasoning or effort. This is a change in the evidentiary meaning of an observable artifact. Signaling theory provides a way to describe the dependence of a signal's interpretation on the processes and incentives that generate it, without transferring the complete structure of a labor-market signaling model to scholarly work [21].

The relevant loss of informativeness is conditional. A formally checked proof can remain unusually informative about the correctness of a result, although its existence alone reveals less about which participant supplied each step. A complex scientific dataset may remain costly to obtain even when its accompanying text is inexpensive to generate. An open-source contribution accompanied by reproducible tests, design explanations, and responsive maintenance may provide richer evidence than an unexamined manuscript with comparable apparent sophistication. Different output signals can thus become more or less diagnostic at different rates. An adequate account should identify the evaluative question for which a signal is used: artifact validity, attributable contribution, professional competence, future reliability, or willingness to cooperate.

The phrase *trust scarcity* similarly requires qualification. Trust is relational and context-dependent, and its supply cannot be represented as a fixed stock independent of evidence or institutional design. What may become scarce in a congested environment is the opportunity for a producer to undergo sufficiently thorough assessment for a warranted trust judgment to form. Such scarcity concerns the time, access, and attention required for informed assessment. A system may contain numerous trustworthy producers while lacking the capacity to recognize them. Conversely, an efficient screening institution may expose unreliable submissions without becoming less capable of trust in general. These possibilities make evaluation design central to the problem.

It follows that an increase in output quantity has at least three analytically distinct potential consequences: greater availability of useful contributions, higher aggregate verification demand, and altered signals about producers. The consequences can move in different directions. A new computational

aid might increase the number of correctly verified results while also increasing the burden on a limited group of evaluators. A similar aid might produce a large number of poorly checked claims that legitimately receive skeptical treatment. In another setting, standardized verification could expand the effective evaluation capacity faster than artifact supply. The output-inflation hypothesis applies only where evidence indicates that signal volume and evaluation constraints interact to reduce the probability or informativeness of substantive recognition.

4.4 Portfolio Inference and Selective Examination

Recognition of a producer differs from certification of a single artifact. Hiring committees, potential collaborators, funding bodies, and readers often encounter portfolios whose scale exceeds the time available for detailed examination. They must select what to inspect and how to infer competence from incomplete samples. The portfolio may contain independent projects, revisions of shared underlying work, exploratory drafts, rigorously verified findings, and artifacts whose apparent novelty arises from changes of presentation. A count without a description of these relationships provides little basis for judging either the diversity or the quality of the underlying contribution.

A hypothetical comparison helps isolate the mechanism. Consider two otherwise comparable researchers whose relevant competence and standards of checking remain constant. One releases three carefully selected artifacts, while the other uses improved tools to release substantially more artifacts, including additional working demonstrations and elaborations. An evaluator given limited time may examine roughly the same number of items from each portfolio. The larger portfolio then creates a more difficult selection problem, despite having been produced by a person with unchanged competence. The evaluator may infer exceptional breadth, suspect duplicated work, request an explanation of AI involvement, or simply defer evaluation. None of these outcomes follows necessarily from portfolio size. The example shows how producer competence, artifact supply, attention constraints, and evaluation rules can vary independently.

To make selection explicit, let $I_{i,o,t}$ equal one when observer o gives artifact i a substantive examination during period t , and zero otherwise. A general selection representation is given in Equation 4.

$$\Pr(I_{i,o,t} = 1) = \pi_o(x_i, n_{a,t}, B_t, z_{o,t}), \quad (4)$$

where x_i denotes observable artifact features, $n_{a,t}$ the size and composition of the producer’s available portfolio, B_t pending evaluation workload, and $z_{o,t}$ the observer’s available resources and institutional conditions. The function π_o is unspecified and must be estimated or otherwise justified for a concrete evaluation system. Equation 4 permits positive, negative, and nonmonotone responses to portfolio expansion. Its analytical purpose is to show where a production-related variable may enter the selection process, not to assign an intrinsic penalty to diligence.

Selection can influence subsequent assessments even when the evaluative rule applied to a fully examined artifact is sound. A producer may have excellent work that is never sampled. An evaluator may draw a broad conclusion from an unrepresentative item. The producer may respond by highlighting a smaller selection of works, providing navigable evidence of a project’s developmental history, or presenting independent validation. Each response modifies the encounter between portfolio and observer rather than the underlying competence that gave rise to the portfolio. A particularly important institutional question is whether the burden of making work examinable falls exclusively on the producer or is shared through accessible evaluation infrastructure.

Institutional heuristics can also be rational under constrained information while introducing systematic errors. Using demonstrated relevance to the institution’s goals can reduce wasted examination time; allowing an applicant to select representative work may improve tractability. Strong reliance on institutional pedigree may sometimes proxy for prior independent checking, but it can also obscure valid work from less-connected producers. Automated triage can identify duplicate text, missing artifacts, or test failures, although its error rates and domain limitations require scrutiny. These practices should be evaluated in relation to the decision being made, the evidence available, the distribution of error costs, and the availability of appeal or substantive reassessment.

Table 3: Proposed mechanisms linking output expansion to the conditions of institutional evaluation.

Mechanism	Proposed evaluative effect	Necessary distinction
Incoming workload growth	Fewer resources for detailed examination under fixed effective capacity	Workload compared with artifact count
Signal-cost change	Altered inference from polished output to producer competence	Artifact quality compared with contribution attribution
Portfolio sampling	Selective scrutiny and dependence on screening rules	Examination probability compared with assessment conditional on examination
Validation support	Lower costs of verifying eligible classes of claims	Technical checks compared with domain-wide epistemic certification
Institutional adaptation	Changes in capacity, screening, and accountability	Context-specific responses compared with universal deterioration

Table 3 summarizes the mechanisms addressed in this section. The table distinguishes alterations in attention and interpretation from claims about the objective quality of knowledge.

4.5 Rival Mechanisms and Boundary Cases

A proposed connection between output expansion and diminished recognition must be evaluated against competing explanations. Increased submission volume may coincide with a change in the average quality of submitted work; if quality falls, a lower average acceptance rate or stricter screening can reflect a justified response. Changes in the composition of submitting populations may also matter. The arrival of new producers, shifts in research fields, altered submission incentives, and variations in the maturity of research programs can change the observed distribution of outcomes. A comparison of recognition before and after adoption of AI assistance that ignores these changes risks attributing ordinary compositional effects to an institutional recognition paradox.

The temporal location of evaluation provides a second set of alternatives. Early drafts may enter repositories before substantial checking, while final products undergo independent review at later stages. An apparent proliferation of research artifacts can therefore reflect a different publication sequence rather than an increase in the number of results presented as verified knowledge. Similarly, visible portfolios can count multiple versions of one evolving artifact as separate outputs. Empirical study requires consistent definitions of artifact identity, revision, and certification status. The trajectory of a work may become more visible through publication practices that reveal intermediate research stages, even where final verified output changes little.

A third set of alternatives concerns expanded evaluation capacity. Automated testing can lower the cost of checking some classes of software behavior; machine-assisted mathematical verification can establish claims within well-specified formal systems; improved metadata and search can reduce the effort needed to identify relevant contributions. Rigorous evaluation still requires careful specification of test conditions, assumptions, and the relation between formal results and intended real-world claims. The National Academies’ treatment of reproducibility and replicability is relevant to these differentiated standards of assessment [17]. When institutional investment increases effective review capacity or improves selection accuracy, higher output need not diminish the probability of warranted recognition.

A fourth alternative involves genuine improvements in the public availability of expertise and knowledge. Researchers using AI may reach collaborators or audiences who were previously unable to inspect their work, particularly when new tools improve documentation, translation, or executable reproducibility. At the field level, a study of broadly defined AI-augmented research reports associations with expanded individual scientific impact alongside collective concentration of research attention [12]. These findings do not establish the mechanisms of trust formation examined in the present paper. They do, however, caution against the assumption that AI-mediated productivity generally reduces recognition. Depending on organizational conditions and evaluative practices, expanded production may coincide with increased recognition, unchanged recognition, or diminished recognition.

Finally, recognition itself has multiple outcomes. A manuscript may receive scholarly criticism without receiving broad attention, a researcher may receive professional trust without institutional appointment, and an artifact may be adopted without a compensating return to its originator. Evaluation congestion concerns the selection and examination of evidence, whereas competence attribution and subsequent material return require additional mechanisms. The paper therefore treats a recognition inversion as a contingent result whose explanation requires holding relevant quality and competence variables fixed, specifying what the evaluator observed, and examining changes in the assessment process. The necessary distinctions will be developed in Sections 5–9.

The central result is a distinction between the amount submitted and the opportunity for substantive examination. When workload expands without comparable increases in effective review capacity, some artifacts may receive less attention; screening rules and the composition of submissions determine how that burden falls. This mechanism alone establishes neither deficient competence nor a negative judgment by an informed evaluator. Section 5 examines the additional evidence required to attribute a completed artifact to particular human capabilities and decisions.

5 Attributional Uncertainty and Competence Inference

Evaluative access and evaluative interpretation are distinct problems. Section 4 examined whether an artifact receives substantive scrutiny; this section asks what that scrutiny can establish about the person who produced it. A finished artifact may support a claim about correctness without revealing which tasks involved human judgment, computational assistance, or independent validation. The discussion therefore separates artifact evidence, contribution roles, task-specific competence, justified verification, and the costs of provenance. These distinctions specify the attributional channel that Section 6 brings into the recognition-inversion hypothesis.

5.1 The Evidentiary Scope of Finished Artifacts

An intellectual artifact is an object available for examination under particular procedures: a manuscript, mathematical derivation, software program, dataset, model, or experimental report. Its observable properties can supply evidence about specific claims. A program may execute successfully on a documented test suite; a proof may satisfy the rules of a formal system; an empirical estimate may follow from a reproducible analysis under stated assumptions. These are consequential observations about the artifact. Their evidentiary reach depends on the standards and coverage of the examination performed. Successful execution on selected inputs does not establish correct behavior under every input, and a formally checked derivation establishes a result relative to the axioms, specifications, and trusted computational components employed. Such limits are ordinary features of evaluation; AI mediation changes the relevance of distinguishing these different questions, rather than creating the distinction itself.

A judgment about a producer requires further inferential steps. Evaluators may want to know whether a person formulated the underlying question, selected the method, recognized a limitation, diagnosed errors, supervised a computational process, or could perform relevant future work. The same finished artifact can be compatible with different combinations of these activities. An accurate scientific software package might reflect extensive human mathematical design and largely automated code generation; substantial manual coding and automated testing; a collaborative team with specialized responsibilities; or a high degree of automated generation followed by limited checking. These configurations can differ greatly in the evidence they provide about the capacities of a particular person. The artifact can possess scientific or practical value across more than one configuration, while the inference from artifact to individual competence varies.

The distinction can be represented without assuming a measurable universal quantity of ability. Let x denote an observed artifact, z a possible production history containing the actions and dependencies relevant to its creation, and F the relation through which a specified history gives rise to an artifact under the conditions of production. The set of histories observationally compatible with the finished artifact is

defined in Equation 5.

$$\mathcal{H}(x) = \{z \in \mathcal{Z} : F(z) = x\}. \quad (5)$$

The domain $\mathcal{Z}$ and the criterion of artifact equivalence are task-dependent. For a document, equality may denote the same final text; for software, it may denote the same submitted code or the same behavior on a test collection. This is a structural representation of an inverse problem, rather than an empirically calibrated model. When multiple production histories remain compatible with the evidence, the final artifact alone may leave features of the production process underdetermined. Additional records or direct assessments can distinguish some histories, but the mere multiplicity of compatible histories does not imply that all histories are equally probable.

This analysis also clarifies the limits of a historical comparison based on effort. A finished artifact previously may have carried information about the time or training required to produce it because certain production methods were prevalent. A broader set of feasible workflows changes the relationship between the same observable artifact and those unobserved inputs. Evaluators who continue to associate length, fluency, or code volume with a particular amount of human labor may be using a convention whose validity depends on production conditions that are no longer stable. Equally, it would be a mistake to infer that an artifact produced rapidly contains little human intellectual direction. The relevant inquiry concerns the evidence connecting a particular output to a particular claim about capability. Signaling theory provides a general framework for considering how observable characteristics operate when important producer attributes remain unobserved; its application to AI-mediated scholarly work is an analytical extension, rather than a direct empirical result of the original labor-market model [21].

An evaluative decision should therefore specify its object. A repository assessed for public reuse calls for evidence of reliability, documentation, licensing, and maintenance. A research paper assessed for publication calls for evaluation of arguments, methods, evidence, and novelty. A person assessed for employment may also need to demonstrate judgment, adaptability, and the capacity to explain or develop the work under changing conditions. Conflating these tasks can generate both unwarranted exclusion and unwarranted endorsement. A high-quality artifact may remain useful despite uncertain individual authorship contributions; a highly capable individual may submit a preliminary artifact that is not yet reliable. Neither case is adequately assessed by a single quantity of output.

Attributional ambiguity creates a further distinction between *uncertainty* and *miscalibration*. If a finished artifact is equally compatible with multiple contribution histories, an evaluator may reasonably withhold a strong conclusion about unaided performance. An adverse competence inference becomes epistemically questionable when its evidential weight is disproportionate to a specified decision, available corroboration, or the likelihood of alternative production histories. Holding the producer's actual competence constant in a hypothetical scenario does not itself establish that an observer's posterior inference is irrational: a rational observer can update incorrectly in an individual case while remaining well calibrated across an appropriate reference population. Any claim of unjustified discount consequently requires a defensible benchmark for the information and evaluation task, rather than a comparison with the unobservable truth of one illustrative individual.

5.2 Configurations of Human and Computational Contribution

AI-mediated knowledge production can redistribute tasks across humans, computational systems, and supporting organizations. The role of a model in generating text, code, a design proposal, or a data transformation does not determine who framed the problem, selected valid assumptions, supplied data, authorized publication, or accepted responsibility for errors. Table 4 sets out several evaluation objects and the limits of possible supporting evidence. These activities are different forms of participation and require evidence appropriate to their character. Contribution analysis consequently benefits from a vocabulary richer than a binary classification of work as human-made or AI-made.

The Contributor Roles Taxonomy (CRediT), formalized as ANSI/NISO Z39.104-2022, differentiates roles including conceptualization, methodology, formal analysis, software, validation, resources, and writing [18]. It was developed to describe contributorship in scholarly work; it does not itself determine

authorship status. The taxonomy offers a useful reference point for the kinds of human contributions that may remain obscured by a final artifact. A person may provide the scientific conception and mathematical constraints while a system generates the initial code. Another person may contribute through independent validation, critical interpretation of the results, or the maintenance infrastructure that makes the artifact usable. Describing these activities can improve the clarity of evaluation, provided that roles are supported by relevant evidence and are not treated as mechanically interchangeable.

Several configurations illustrate the analytical problem. In a *human-directed implementation* configuration, a researcher defines a scientific method, constrains computational requirements, and evaluates generated implementations. The final amount of machine-generated code may be large even where substantive human expertise is essential. In a *machine-proposed research* configuration, a model suggests a hypothesis or proof strategy that a human investigator interprets and attempts to establish. Here, the claims of originality and human intellectual contribution depend on activities beyond writing the final explanation. In a *human-validated automated workflow*, computational systems may execute much of the pipeline while humans specify validity criteria, inspect anomalies, and decide which results warrant disclosure. Conversely, in an *unexamined generation* configuration, fluent outputs can be produced and circulated with limited evidence of successful human validation. Each configuration merits an assessment of the particular results and responsibilities involved. They cannot responsibly be ranked solely by the proportion of text or code emitted by a model.

A useful analytical distinction separates *causal contribution*, *epistemic contribution*, and *responsibility*. Causal contribution concerns actions or resources without which a particular artifact might not have arisen in its observed form. Epistemic contribution concerns the formulation, testing, interpretation, or warranted acceptance of claims. Responsibility concerns who can be called upon to justify choices, correct errors, disclose limitations, and respond to institutional or ethical obligations. The three categories overlap without being identical. An external computing service may be causally indispensable, yet an organization or author can retain responsibility for the reported scientific conclusions. Human oversight may be nominal in one workflow and technically substantive in another. These differences require examination rather than inference from a final file alone.

Publication guidance already recognizes aspects of this separation. The International Committee of Medical Journal Editors requires disclosure of AI-assisted technology in submitted work and assigns responsibility for the accuracy, integrity, and originality of published content to human authors [13]. The Committee on Publication Ethics similarly states that AI tools do not meet the requirements of accountable authorship and calls for transparent disclosure of relevant uses [5]. These statements apply to their respective publication-policy contexts; they do not provide a complete causal theory of contribution or settle every cross-disciplinary authorship question. Their significance for the present argument lies in demonstrating that tool participation and responsible authorship can be considered separately. Evaluative institutions can recognize consequential computational assistance while identifying humans or organizations answerable for the claims being advanced.

Table 4 identifies selected kinds of evidence and the inferential limits they carry. The table is a proposed analytical aid, not an instrument validated for measuring competence.

The distinctions in Table 4 also show why contribution transparency has to be evaluated according to purpose. A role statement can reduce uncertainty about declared responsibilities without proving that every claim is correct. A comprehensive sequence of commits can document modifications without establishing who conceived the underlying method. Direct observation of technical reasoning can provide stronger evidence about an individual's understanding while remaining sensitive to the question posed, the interviewer's expertise, and the support permitted during examination. Robust evaluation often combines sources of evidence instead of expecting one document or metric to carry the full inferential burden.

5.3 Competence, Understanding, and Task-Specific Demonstration

Competence is multidimensional and context-dependent. The ability to formulate a promising problem, evaluate scientific evidence, design a reliable algorithm, interpret a proof, coordinate contributors, and

Table 4: Evidence for different evaluation objects in AI-mediated intellectual production.

Evaluation object	Potential evidence	Principal inferential limitation
Artifact quality	Replication, formal checking, expert examination, tests	Coverage depends on assumptions, specification, and examination scope
Conceptual contribution	Design notes, problem formulation, methodological explanation	Retrospective narratives may be incomplete or strategically selected
Implementation contribution	Version history, reviewed changes, reproducible build records	Recorded edits alone do not identify intellectual authorship or comprehension
Individual competence	Explanation, unfamiliar problem-solving, error diagnosis	Performance is task-specific and affected by examination conditions
Human accountability	Disclosure, named responsibility, documented review and correction	Responsibility assignments do not independently prove correctness
Sustained reliability	Maintenance history, correction practices, repeated independent outcomes	Prior performance may have limited transportability to new contexts

maintain software over time may be distributed unevenly across persons and teams. Evaluators also differ in the attributes relevant to their decisions. A laboratory recruiting a research programmer may require evidence of debugging and technical maintenance; a theory group may place greater emphasis on conceptual judgment, mathematical explanation, and identifying unsupported assumptions. A general declaration that an artifact proves or disproves competence ignores this dependence on tasks and roles.

For a specified assessment, let θ denote the competence attributes relevant to that task, x the finished artifact, d additional evidence about production and validation, and m the assessment conditions. Equation 6 expresses the dependence of a possible probabilistic inference on these observations.

$$p_o(\theta \mid x, d, m) \propto p_o(x, d \mid \theta, m) p_o(\theta \mid m). \quad (6)$$

The subscript o denotes an evaluator-specific assessment model, and the expression assumes the probability terms are defined and the evidence can be represented within the stated conditioning variables. It does not propose that real evaluators use calibrated Bayesian inference, nor that competence can be observed through a single latent score. Its purpose is to identify the role of additional evidence: judgments derived from x alone can differ from judgments incorporating relevant information about production and examination. A more detailed cognitive model of how human observers form these judgments belongs to the separate investigation of trajectory recognition, rather than to the present institutional account.

Demonstrable understanding can be explored through targeted, proportionate tasks. A person might be invited to explain why an algorithm satisfies a stated invariant, identify circumstances in which a simulation becomes invalid, revise a proof after an assumption changes, or diagnose a model failure on a previously unseen example. These activities provide evidence about particular skills when the prompts, allowed tools, and criteria are matched to the capability being assessed. They are neither infallible tests of general scientific ability nor appropriate for every evaluation setting. Performance may be affected by language, disability accommodations, examination anxiety, familiarity with the testing format, and access to the computational environment. Assessment design must therefore consider both the validity of the elicited evidence and the costs imposed on participants.

Repeated and independently checked performance can also be informative. A record of identified and corrected errors may offer evidence of research responsibility that a polished final submission conceals. A successfully replicated experiment may support claims about a method’s reliability while providing limited evidence about who first conceived it. The National Academies’ treatment of reproducibility and replicability underscores the need to specify what an examination actually establishes [17]. For the present paper, the relevant implication is methodological: artifact-level verification, contribution attribution, and competence assessment can each require different evidence and different procedures. An institution that evaluates one object should state whether it is extending the resulting judgment to another.

There is an additional asymmetry in access to demonstration opportunities. Some producers receive structured interviews, research presentations, code reviews, or collaborative trial periods, while others

are filtered out through preliminary signals and never reach those stages. In settings where a polished artifact has become weak evidence of competence, the availability of direct interaction may acquire greater importance. Access to such interaction can nevertheless depend on professional networks, institutional status, or an evaluator’s discretionary allocation of time. These relations remain a preliminary institutional consideration in this paper. Their structural dynamics and cognitive consequences require the dedicated analyses of heterogeneous relational events and observer cognition in the accompanying research programme.

5.4 Warranted Verification and Unsupported Credibility Discounts

Uncertainty about the history of an artifact justifies attention to the relationship between available evidence and the conclusion an evaluator is being asked to endorse. Its practical consequences vary by domain and stakes. A claim informing a medical decision, a mathematical proof supporting further results, and a preliminary exploratory essay require different kinds and intensities of scrutiny. The presence of AI assistance may be relevant when the methods used introduce known or suspected error patterns, prevent adequate replication, or obscure responsibility. A request for technical explanation, independent verification, or clear disclosure can therefore be justified even when the submitted artifact appears coherent and complete.

An unsupported credibility discount arises under a different configuration: an evaluator lowers an assessment of artifact quality or individual competence chiefly on the basis of generalized assumptions about AI mediation, without sufficient evidence that the assumptions apply to the work or person examined. An illustrative case involves two hypothetical computational results with equivalent independently verified behavior. One result is accompanied by an accurate disclosure of model-assisted code generation, while the other supplies only the final implementation. If evaluators assign systematically lower technical quality to the disclosed result after receiving the same verification evidence, the difference would warrant investigation. It would not, by itself, establish unfair treatment of an individual producer, because additional conditions concerning the evaluation purpose and the available evidence would still matter. The example clarifies a potential mechanism, not a documented empirical effect.

The converse error is also possible. A transparent account of substantial AI assistance cannot substitute for testing an artifact’s claims, and a detailed explanation of a production history does not automatically demonstrate domain expertise. Accuracy, originality, lawful use of sources, and ethical responsibility remain legitimate objects of assessment. ICMJE guidance specifically highlights the risk that AI-generated content can appear authoritative while being incorrect, incomplete, or biased [13]. These concerns justify evidence-sensitive evaluation. They do not establish that the expected quality of every AI-assisted artifact is low, or that extensive use of a system is incompatible with demanding human intellectual work.

A useful criterion is *evidentiary proportionality*. The evidence required for a decision should be connected to the attributes actually at issue, the stakes of error, and the information already established. A technical evaluator may appropriately request a proof of correctness for a safety-critical routine; a general portfolio screen may have sufficient grounds to ask for one independently explained representative contribution. By contrast, requiring exhaustive reconstruction of every prompt, intermediate draft, and exploratory branch may produce heavy costs while providing little additional decision-relevant information. Proportionality also permits contestation: an applicant or author should, where feasible, be able to clarify a disputed inference, supply missing validation, or correct a mistaken attribution. Such procedures are proposed institutional responses whose effectiveness requires testing; they are not universal requirements established by the cited publication guidance.

The distinction between justified scrutiny and unwarranted discount becomes especially important under congestion. When evaluator time is scarce, a shorthand inference from unusual productivity to likely unreliability can appear administratively efficient. It may sometimes reflect a reasonable statistical concern about the distribution of unvalidated submissions. It may also exclude genuine contributions before their supporting evidence is examined. Assessing these competing explanations requires data on

evaluator selection, accuracy, the prevalence of relevant errors, and the costs of verification. Section 6 will examine the conditions under which such heuristics could produce an inverted relationship between additional output and recognition.

5.5 Provenance, Disclosure, and the Distribution of Evaluation Costs

Provenance concerns the documented origins, transformations, participants, and dependencies of an artifact. It can support reproducibility, responsibility assignment, and the reconstruction of decisions. The World Wide Web Consortium's PROV-O recommendation offers a general representational vocabulary connecting entities, activities, and agents, including relations of derivation and attribution [15]. This vocabulary can describe portions of a computational or scholarly history in interoperable form. The existence of a provenance statement, however, is distinct from its truthfulness, completeness, and probative relevance. A system can record a dependency accurately while leaving the scientific adequacy of its output unexamined. Provenance infrastructure accordingly improves the possibilities of examination under specified conditions; it cannot replace substantive epistemic evaluation.

The level of recording also matters. At one end, a brief disclosure can identify the model class, general purpose, and responsible human contributors. At the other, detailed records may include input versions, execution settings, prompts, tests, rejected alternatives, code changes, computational environments, and data-access arrangements. The latter can enable unusually fine-grained reconstruction of some decisions, although many histories remain incomplete even under extensive logging. The additional evidence may be valuable for high-stakes scientific claims and methodologically complex workflows. It may be less useful in routine applications where the relevant validity question can be settled through direct testing or independent review. A well-designed provenance requirement therefore states which relations are necessary to establish a particular claim and which evidentiary gaps remain after documentation.

Transparency can also impose costs or conflicts. Detailed logs may contain confidential research material, personal data, unpublished hypotheses, proprietary code, third-party intellectual property, or security-sensitive information. Retaining and organizing these records consumes computational resources and human time. A small independent producer and a well-resourced laboratory may face markedly different burdens when asked to supply the same documentary package. Uniform requirements can consequently create unequal access to evaluation even when stated in formally identical terms. Institutions may need alternatives such as restricted-access audits, versioned methodological summaries, selected reproducibility artifacts, independently witnessed demonstrations, and contestable explanations of why particular evidence is insufficient. The merits of these options depend on domain, feasibility, confidentiality obligations, and the stakes of the decision.

Provenance requirements also generate a potential regress: demonstrating that a record is authentic may call for evidence about how the record was created, while establishing the authenticity of that evidence may require further checks. Practical systems resolve this problem through bounded trust assumptions, independent audits, tamper-evident mechanisms where appropriate, and institutionally specified standards of sufficiency. Recording every possible trace is neither technically feasible nor epistemically necessary. The appropriate stopping point depends on what is being evaluated and the consequences of error. A source of additional concern is that highly polished provenance narratives may themselves be produced with computational assistance; the aesthetic coherence of a process description is not proof of its correspondence to events.

These considerations yield a conditional institutional proposition. When production methods diversify, a credible evaluation system may benefit from explicit descriptions of contribution roles, targeted tests of competence, and proportionate access to provenance. The effectiveness of those measures depends on the quality of evidence, the evaluator's resources, and opportunities to contest errors. They may support both more reliable validation and fairer participation, but they can also create new exclusionary costs if designed without attention to practical access. The response to attributional uncertainty therefore cannot be restricted to demands placed on individual creators; institutions also face responsibilities

concerning the design, financing, and correction of evaluative procedures.

The attribution problem concerns what an examined work demonstrates about its producer. Evidence of execution or validity can coexist with uncertainty about task allocation, independent judgment, or responsibility. Task-specific demonstrations and proportionate provenance may reduce that uncertainty, while generalized documentation demands can transfer substantial costs to producers. Section 6 considers when these interpretive conditions contribute to an adverse response to reported productivity.

6 Recognition Inversion in AI-Mediated Production

This section formulates recognition inversion as a conditional response to observable productivity under AI-mediated production. Quantity can serve as evidence of persistence or competence when the production process is sufficiently understood; changes in task costs can alter that evidentiary relation. The analysis examines the established signaling convention, the information lost or gained when production methods change, the distinct dimensions of recognition, and counterexamples to an adverse response. Studies of AI labels and disclosure supply adjacent evidence about evaluation, although they do not estimate recognition inversion in scientific careers. The central empirical question is whether a negative productivity response reflects warranted inference, restricted examination, or a poorly calibrated judgment.

6.1 Conventional Productivity Signals

Institutions frequently encounter producers through incomplete records: published articles, submitted manuscripts, repositories, research statements, presentations, work samples, and descriptions of completed projects. The number and timing of these artifacts are readily observable even when the underlying judgments, discarded alternatives, and production arrangements are not. In some settings, a sustained record of completed work can reasonably indicate persistence, technical familiarity, the ability to organize research, or access to effective production arrangements. These interpretations need not be mistaken. They become problematic when the same visible features are treated as interchangeable evidence of distinct and unobserved attributes, or when historically situated expectations about the work required for an output are treated as fixed.

A productivity signal can be meaningful for several reasons. Repeatedly completed tasks may demonstrate reliable execution across changing circumstances. A record of independently validated results can show the capacity to notice and correct errors. Work across several related applications may reveal an ability to adapt common methods. In these cases the evidential content of the portfolio depends on the tasks undertaken, the constraints encountered, the extent of external validation, and the evaluator's question. Mere length or count is an incomplete summary of this content. A large set of closely related artifacts may reveal considerably less about methodological breadth than a smaller set produced under different and independently demanding conditions.

Signaling theory offers an established general vocabulary for interpreting visible characteristics when other relevant attributes remain difficult to observe [21]. Within its original domain, a signal's informational role depends on the structure of incentives and the costs borne by different types of senders. The present application to scholarly and software production is an extension requiring additional assumptions. Neither an observed number of publications nor the time reportedly spent producing them automatically constitutes a separating signal of the underlying intellectual competence that an evaluator seeks to assess. Where publication depends on gatekeeping, resources, field-specific conventions, and collaborators, productivity also reflects organizational opportunities and conditions of access.

The conventional interpretation of diligence is similarly heterogeneous. Diligence may refer to sustained effort, repeated verification, disciplined attention to error, or the persistence needed to complete a difficult problem. Its observable correlates are imperfect: a methodologically careful project may yield one small artifact over a long interval, while an efficient computational workflow may generate several useful outputs. An evaluator can reasonably commend diligent work without treating a high artifact count

as definitive evidence of unusually great effort. Conversely, apparent rapidity alone does not establish the absence of difficult human judgment. The appropriate inferential question is which capacities and contributions the available record actually distinguishes from plausible alternatives.

These distinctions become consequential when comparing a producer's record across technical regimes. If the same individual acquires tools that substantially reduce implementation or writing time while continuing to perform the relevant conceptual and verification work, the ratio between artifacts and observable human effort changes. The producer's competence need not have changed. The evidentiary meaning of the output count can nevertheless change for an observer who is familiar with earlier production conditions. An institution may recognize this transition and recalibrate its evaluation. It may also retain shortcuts through which unusual quantity is taken as either outstanding competence or an indication that too little human intellectual work could have occurred. Both shortcuts require examination rather than automatic acceptance.

6.2 Signal-Cost Changes and Evaluative Interpretation

AI assistance can reduce the costs of drafting, routine coding, translation, documentation, search, and preliminary variation across solutions, while leaving other costs comparatively high. Problem selection, establishing an adequate scientific specification, acquiring relevant evidence, independent validation, interpretation of limitations, and responsibility for dissemination may remain demanding. The resulting production history consists of tasks with different technologies, uncertainties, and participants. Section 3 described these transformations at the level of production mechanisms and relations; Section 5 showed that an artifact may be compatible with different contribution histories. Their combined implication is that the amount of output becomes less diagnostic of any single component of individual ability when production arrangements diversify.

A decline in the cost of producing a fluent document need not entail a comparable decline in the cost of producing a valid scientific result. Evaluation may become more demanding precisely because the observable presentation is less informative about the work performed to verify a claim. However, a further inference from reduced generation cost to reduced quality is unwarranted without evidence about the relevant content. The cost of generating a plausible answer and the reliability of a checked answer are different objects. For an evaluator assessing a manuscript, the validity of arguments and evidence remains central; for an evaluator assessing a person, direct demonstrations of understanding and contribution may provide additional information. A shift in the relation between effort and output modifies which observations are informative, rather than assigning a uniform quality level to AI-mediated products.

An accurate disclosure of AI use can make this inferential difficulty salient. Some receivers may interpret the disclosure as helpful information about the production process; others may infer extensive automation even where a human contributed conception, method, and validation. Adjacent experimental evidence demonstrates that the interpretation of AI-source information can alter judgments under particular conditions. In two preregistered experiments involving headlines, Altay and Gilardi found that an AI-generation label reduced perceived accuracy and willingness to share, including for materials whose veracity and actual provenance did not justify an across-the-board penalty [1]. The study concerns news-headline judgments, and cannot establish an analogous penalty in scholarly competence evaluations. It nevertheless supports the possibility that source labels alter evaluations independently of some artifact properties.

A further series of sixteen preregistered experiments on creative writing found lower evaluations when participants believed the writing had been generated by AI or with AI assistance; perceived authenticity was identified as a mediating factor in that context [20]. These findings concern judgments of creative writing, which can legitimately involve aesthetic expectations and perceptions of authorship that differ from the standards used for mathematical proofs, scientific software, or empirical research. Their relevance to the present theoretical argument is correspondingly limited: they illustrate that source attribution may change how an otherwise comparable artifact is received, and motivate studying whether

similar mechanisms influence assessments of human research competence.

There is also important contrary evidence. In a survey experiment on evidence-based policy messages, Gallegos and colleagues reported no statistically significant effect of AI-versus-human authorship labels on attitude change, perceived accuracy, or sharing intention in their tested setting [7]. The null result prevents a simple generalization from some disclosure penalties to all AI-labeled artifacts or all dimensions of audience response. The experiments also differ in task, stimulus, evaluator, and dependent variable. Their proper joint interpretation is that disclosure effects appear context-dependent, and that the more specific claim of recognition inversion under high scholarly productivity requires its own identification strategy.

Signal costs also have distributive implications. The same reduction in some technical production costs is experienced differently by a well-resourced research group and a person conducting research with limited funding. Access to computational tools, data, independent validators, collaborators, and reputable venues remains unequally distributed. An evaluator observing high output may draw inferences from the tools available to a producer, yet those inferences can miss the human work of integrating heterogeneous tasks. A robust model must therefore distinguish the *feasibility* of producing an artifact, the *observability* of its creation, and the *diagnostic value* of the evidence presented. Neither technical acceleration nor AI-source disclosure by itself settles these questions.

6.3 Conditional Recognition Inversion

The recognition paradox becomes more specific when the outcome and comparison are identified. A producer might receive less substantive examination despite producing more eligible outputs. A reviewer might report lower confidence in the producer’s conceptual competence after viewing a high-volume portfolio. An institution might delay endorsement because its procedures cannot accommodate the apparent production pace. These are distinct relationships. An adverse assessment may arise from a genuine difference in evidence, from a defensible allocation decision under scarcity, or from a heuristic that incorrectly substitutes artifact count or production method for competent participation. The existence of an adverse outcome does not, without further analysis, establish the last explanation.

For a deliberately restricted comparison, let n denote the number of reported eligible outputs over a specified interval and let $R_k(n; c, q, z)$ represent the expected value of a named recognition outcome k under a comparison protocol. The parameter c denotes the producer’s relevant underlying competence, q the verified quality of the outputs on which the evaluation is based, and z a collection of specified evaluation conditions, including the evidence available and the task the evaluator is asked to perform. Equation 7 defines an adverse comparison between two feasible portfolio presentations.

$$\Delta_{n_1, n_2} R_k \equiv R_k(n_2; c, q, z) - R_k(n_1; c, q, z) < 0, \quad n_2 > n_1. \quad (7)$$

The comparison is a formal representation of the hypothesis, not a measured relationship. The interpretation of holding q fixed must be specified: it may refer to a representative artifact with independently assessed quality, or to a controlled distribution of validated artifacts, rather than a claim that total portfolio quality is unchanged when new items are added. Likewise, z cannot be assumed identical if changing output necessarily alters disclosure, search costs, or reviewer workload. An empirical design must either hold the relevant conditions stable by construction or estimate how their changes mediate the result. Equation 7 alone does not establish a causal effect of additional production.

The sign of a productivity response also admits competing epistemic interpretations. A negative response can be a rational use of probabilistic evidence if, for the specified skill and reference population, unusually large portfolios are genuinely less diagnostic of that skill after production arrangements change. An empirically observed penalty becomes evidence of systematic miscalibration only when the response is assessed against relevant contribution and quality data, baseline expectations, and a prespecified evaluative purpose. The recognition-inversion hypothesis therefore concerns the direction of an observed conditional response, whereas a separate normative assessment concerns its justification. Neither should be inferred from the other by definition.

Hypothesis 6.1 (Conditional recognition inversion). For some combinations of evaluation purpose, available evidence, observer expectations, and institutional constraints, an increase in reported AI-mediated output is associated with a less favorable competence attribution or other specified recognition outcome despite adequately controlled underlying competence and relevant artifact quality. The direction, size, and persistence of this association are open to empirical determination.

Hypothesis 6.1 specifies a conditional possibility rather than a general expectation. At least two routes to an adverse comparison should be kept distinct. The *examination route* arises when increasing output competes for finite evaluator time. It may reduce the probability that a particular artifact is read carefully even when an evaluator has no negative belief about the producer. Section 4 described the corresponding capacity and selection constraints. An observer who declines to inspect twenty manuscripts because of limited time has made an allocation decision whose relation to competence attribution requires independent measurement. The *attribution route* arises when the reported pace or mode of production changes beliefs about the producer's actual role or capability. These routes can reinforce one another, but their empirical signatures differ. The former can be detected through review selection and delay; the latter requires direct measures of competence inference, explanatory reasoning, or related judgments.

A third route can arise through *institutional threshold rules*. Consider a hypothetical committee using minimum requirements concerning journal standing, independently witnessed demonstrations, or documentary evidence of contribution. A high-output producer may fail to meet a specified verification rule, while a lower-output producer has already accumulated the required endorsements. Such an outcome may reflect the institution's evidentiary design, its resource constraints, or actual differences in validated achievements. It should not automatically be categorized as a biased inversion. A theoretically interesting case exists where equally substantiated contributions are assessed differently because the quantity, appearance, or AI disclosure of one portfolio triggers additional hurdles without a decision-relevant justification. The explanatory burden then concerns the justification and calibration of the evaluation rule.

A possible positive-feedback mechanism warrants careful consideration. Suppose a producer who receives little substantive examination responds by creating more outputs in order to make competence more visible. Additional submissions can increase the cost of examining the portfolio. Evaluators then rely more heavily on selection shortcuts; the producer interprets the absence of response as a reason to produce still more. Such a process could reproduce the adverse conditions it was intended to remedy. This is a *proposed* feedback mechanism, not a diagnosis of any particular creator's situation. Its formal study would require observations across repeated interactions and controls for producer selection, quality, field demand, and institutional resources. The next section examines the resulting choices as a potential tragedy of productive diligence; Section 8 will specify a parsimonious system-level representation.

The magnitude of an apparent inversion also depends on its comparison baseline. A portfolio receiving more total reads but fewer reads per artifact has experienced a per-item decline, not necessarily a decline in aggregate recognition. Likewise, a producer receiving no formal endorsement despite added output has not necessarily lost recognition relative to an earlier baseline. The analysis should identify whether the outcome is absolute, per-artifact, per-unit evaluation cost, relative to peers, or relative to the opportunities that would have existed under a different presentation. Ambiguity on this point can convert ordinary congestion or unchanged recognition into an unsupported claim of a punitive response.

6.4 Disaggregated Recognition Outcomes

Recognition is a sequence of potentially separable processes. An artifact may become visible without being examined; it may be examined without being judged correct; it may be judged valuable without providing decisive evidence of any one person's contribution; and it may be endorsed without generating employment, funding, or other material support. Treating these outcomes as a single confidence score obscures both the mechanisms of inversion and the institution's responsibilities. Table 5 organizes selected dimensions and their observational implications.

A reviewer can assign a favorable artifact-quality score and remain uncertain about the producer's unaided programming skill. That pattern is coherent where assessment of the artifact and assessment

Table 5: Distinct recognition outcomes and interpretations of an adverse productivity association. Each outcome requires its own measurement and comparison basis.

Outcome	Possible observable indicator	Interpretation requiring further analysis
Evaluative access	Invitation to review; reading or testing time	Attention constraints, routing, or exclusion
Artifact assessment	Validity or quality rating after examination	Evidence-based assessment or source-related discount
Competence attribution	Judgment of demonstrated human capability	Inference about contribution, method, and understanding
Institutional endorsement	Acceptance, certification, or professional recommendation	Formal requirements, authority, and evidentiary standards
Relational opportunity	Invitation to discuss or collaborate	Fit, perceived reliability, and accessible networks
Material return	Paid work, research support, or maintenance resources	Recognition, demand, budgets, and conversion arrangements

of unaided skill serve different objectives. An employer may value the person’s ability to direct and validate AI-assisted systems rather than their capacity to recreate every implementation step without tools. Conversely, a position that specifically requires unaided performance may legitimately ask for such demonstration. An appropriately designed evaluation therefore identifies the competence needed for the role before treating the production workflow as favorable or adverse evidence. The same artifact history can support different conclusions under different assessment purposes without implying arbitrary judgment.

The distinction between epistemic confidence and institutional endorsement is equally consequential. A scientist may regard a proof as correct but lack the authority to confer a credential. A hiring committee may accept a candidate’s demonstrated competence yet have no funded position available. A journal may reject a valid result for reasons of scope or duplication. These outcomes concern different institutional decisions, some of which are constrained by resources or organizational mission. The present paper therefore reserves the detailed treatment of resource-return mechanisms for Section 9, while retaining the recognition–return distinction in the analysis of productivity.

This disaggregation also prevents a second measurement error: combining outcomes that can move in opposite directions. Exposure may rise when a producer releases more work, while the fraction of work independently checked falls. An unusual publication pace may increase curiosity in one audience and decrease competence estimates in another. Greater transparency can improve the reliability of quality assessments while placing additional burdens on applicants. For empirical purposes, the dependent variable must represent the phenomenon under investigation, not an arbitrary aggregation of readership, credibility, acceptance, and economic reward. The units and time horizon of comparison must be explicit.

The relationships may also be sequential without being inevitable. Access to examination can enable evidence-based judgments; those judgments may support collaboration; collaboration can produce opportunities for practical use or paid work. Yet each transition can fail for reasons distinct from the previous step. The existence of an artifact with demonstrable value is therefore insufficient to predict downstream returns. Institutions that wish to improve the recognition of valuable work face questions of evaluation resources, task definition, evidence design, and material support. The analysis suggests that these questions should be examined together while retaining their separate empirical statuses.

6.5 Counterexamples and Null Cases

A plausible theory of recognition inversion must explain circumstances in which greater output improves recognition or has no adverse effect. A producer who supplies many independently checked software modules may provide stronger evidence of reliable competence than one who supplies a single preliminary prototype. A researcher whose portfolio contains diverse, technically demanding results may enable

evaluators to estimate capability more accurately. An institution with effective triage and adequately resourced independent review may accommodate additional submissions without lowering per-artifact examination probability. In these cases output quantity can remain an informative positive signal or cease to matter once direct evidence is available. The possibility of such counterexamples is a requirement of the theory's scope, not a defect to be eliminated by broadening the meaning of inversion.

Some unfavorable responses to quantity may also be evidentially justified. If rapid expansion is accompanied by repeated factual errors, inadequate source checks, duplicated claims, or insufficient validation, reviewers may reasonably lower confidence. If an applicant asserts personal performance of tasks demonstrably performed by others, doubts about attribution are similarly warranted. These situations differ from the controlled comparison in Definition 2.1, where relevant competence and quality are held fixed or adequately controlled. A study that fails to measure genuine quality variation might mistake warranted quality-sensitive evaluation for an unsupported discount. Conversely, a study that adjusts away changes in review behavior caused by disclosure might inadvertently remove the mechanism it seeks to estimate. Identification requires a theory of which variables are causes, which are outcomes, and which are confounders.

Institutional adaptation can reduce the reliance on crude output signals. Standardized demonstrations of a relevant method, independent reproducibility checks, meaningful contributor-role descriptions, and clear criteria for representative submissions can provide evidence closer to the actual evaluative object. Such practices may make the number of artifacts less consequential, but they can impose costs and can be unevenly available. A system that demands exhaustive process histories from every producer may protect some evaluations while excluding those without support for documentation and auditing. The appropriate response is consequently an institutional design problem rather than a simple demand that creators personally prove the authenticity of every stage of work.

The evidence reviewed in this section supports a cautious conclusion. Studies of source labels identify settings in which AI attribution changes evaluations and a setting in which tested persuasion outcomes are comparatively insensitive to AI labels [1, 20, 7]. They establish neither a general penalty for scholarly productivity nor a universal incapacity of evaluators to assess AI-mediated work. The hypothesized inversion requires direct research on reported output volume, independently established technical quality, competence attribution, review opportunity, and the institutional conditions under which those judgments are made. Section 11 outlines possible designs and rival explanations.

Recognition inversion is consequently a testable possibility within a wider range of evaluative responses. More output may improve recognition when it supplies additional valid evidence; it may have little effect where review procedures remain informative; and it may produce adverse results through restricted attention or changes in attributional inference. The corresponding producer choices are examined in Section 7. Their costs may arise even without an unjustified credibility penalty.

7 The Tragedy of Productive Diligence

Recognition uncertainty can change how intellectual producers allocate their time even when the underlying work remains technically sound. This section develops the *tragedy of productive diligence* as a conditional conflict among expanding a portfolio, restraining release, documenting contribution, seeking visibility, and sustaining further work. Each choice can provide benefits while imposing opportunity, evaluation, or livelihood costs. The argument uses hypothetical comparisons and resource accounting to identify conditions for such conflicts. It does not require an adverse inference in every institution or an assumption of evaluator misconduct; the recognition-inversion mechanism of Section 6 is one possible contributor to a broader problem of productive sustainability.

7.1 Accelerated Production and Recognition Costs

Accelerated intellectual production can be attractive to an individual researcher, a scientific team, or an organization for reasons independent of reputation. A productive system may enable the examination of

more hypotheses, the implementation of alternative algorithms, the dissemination of teaching materials, or the correction of previously inaccessible technical limitations. Where the resulting artifacts are useful and adequately checked, an increase in throughput can represent a substantive gain. Evidence of productivity improvements in particular AI-assisted occupational tasks illustrates the practical possibility of such gains, although those findings do not establish their magnitude or consequences for independent scientific inquiry [19, 3]. The central question is what happens when output can expand more rapidly than the opportunities to make its significance and provenance intelligible to others.

In an evaluation environment with limited reviewer time, each additional artifact may compete for scarce attention with earlier artifacts from the same producer and with submissions from others. Evaluative congestion, analyzed in Section 4, can reduce the share of work subjected to substantive examination even if the average quality of submitted work is unchanged. Quantity may also change the inferences an evaluator draws about the process that produced a portfolio. A high output rate can provide evidence of effective organization and technical mastery, but it may create attributional uncertainty when the evaluator lacks reliable information about human judgment, computational delegation, validation, and revision. These are distinct mechanisms: congestion concerns the allocation of evaluative capacity, while attribution concerns what an observed artifact can establish about a producer's particular competencies.

Consider a hypothetical researcher who develops a common numerical library and subsequently produces several domain-specific applications. Each application is accompanied by tests, documentation, and an explicit record of limitations. An evaluator with time to inspect the common library and one representative application may be able to identify the researcher's contribution. An evaluator presented only with a long chronological list of releases may encounter considerably more material than can reasonably be checked. The apparent scale of the portfolio is greater in the second presentation, while the evidence relevant to the evaluation question may be less accessible. The difference illustrates a problem of evidentiary organization rather than a defect intrinsic to productivity.

Acceleration can also generate obligations beyond the initial creation of an artifact. Software requires maintenance, dependency updates, security assessments, reproducibility checks, documentation, and responses to reported errors. Scholarly claims invite clarification, correction, and sometimes replication. If the responsible producer has limited support, the number of artifacts in circulation may increase the continuing work required to sustain their credibility and usefulness. The production of further outputs can then compete with the care necessary for previously released work. This conflict becomes especially salient when existing institutions reward visible novelty more readily than maintenance, negative results, or careful correction. The relevant institutional incentives vary by domain and must be investigated rather than presumed.

A further cost is the possible pressure to demonstrate understanding of the very outputs that accelerated methods made feasible. An individual who used AI to assist implementation may have to explain the original problem definition, the selection among alternatives, error analysis, and the boundaries of the generated implementation. Such requests can be proportionate when they relate directly to a role or a technical claim. They can become costly when repeated across different evaluators, requested without a clear assessment purpose, or applied more extensively to AI-assisted work than to comparable work completed with other forms of assistance. The dilemma emerges when additional production simultaneously expands the portfolio and the burden of establishing its meaning.

These costs are neither uniform nor necessarily cumulative. Some producers benefit from modular software designs, standardized experimental pipelines, qualified collaborators, and automated checks that lower subsequent evaluation and maintenance costs. Some evaluators learn to interpret AI-assisted work without applying a generalized discount. The argument therefore identifies conditions of possible adverse feedback, rather than a stable negative relationship between output quantity and trust. The cases of successful large-scale, independently checked production are essential tests of the theory.

7.2 Restrained Production and Foregone Possibilities

If a rapidly expanding portfolio can face congestion and attribution problems, a plausible individual response is to restrict the number of visible outputs. Producers may prioritize a few representative results, delay release until a stronger evidentiary package is available, or suspend new projects while earlier work receives review. Such strategies can improve communication and reduce maintenance costs. Yet restraint may also sacrifice opportunities that would otherwise arise from productive capacity. The potential sacrifice is particularly relevant to research whose value depends on the availability of a reusable method or shared infrastructure to other participants.

Restraint entails several analytically distinct costs. A delayed result may postpone an opportunity for external correction or independent development. A withheld open implementation may temporarily deprive other researchers of a useful instrument. A producer may forego the chance to pursue a time-sensitive question or test an alternative before resources and collaborators become unavailable. Where publication, disclosure, or intellectual-property rules create path dependence, the timing of release may also affect the options available to participants; the direction of that effect depends on the rules governing the particular contribution. Early disclosure can support public scrutiny and documented intellectual priority, yet it can also prejudice patent claims or disclose information that cannot ethically or legally be shared. A theory of productive diligence should therefore avoid prescribing immediate public release as a universal remedy.

The consequence of restrained production is most evident when the producer's active portfolio includes complementary projects. A technical method constructed for one inquiry may later make several other lines of inquiry possible. Postponing the shared method can delay all of those applications. Conversely, attempting every application simultaneously can undermine testing, maintenance, and comprehensibility. The relevant choice concerns allocation of finite effort across exploration, implementation, verification, communication, and institutional engagement. More outputs need not always produce more usable knowledge, but fewer outputs need not always be the prudent choice either.

A neutral thought experiment helps specify the conflict. Two otherwise comparable researchers have identical capabilities, access to the same AI tools, and a set of proposals with uncertain but potentially valuable applications. Researcher A concentrates on one validated artifact and devotes the remaining time to external explanation and evaluation. Researcher B completes additional validated artifacts and devotes less time to contact with potential evaluators. The first trajectory may improve the intelligibility of one contribution while delaying other useful work. The second may increase the available stock of knowledge while leaving important contributions unexamined. Which trajectory better supports scientific value or material sustainability cannot be determined from output counts alone. It depends on the structure of external demand, review capacity, the complementarity of the proposals, and the consequences of delay.

This thought experiment also reveals the limitation of a purely reputational solution. If the sole response to evaluative overload is to require every individual producer to reduce output, a collective evaluation-capacity problem is transferred to the individuals who have become more capable of producing. Such restraint could be sensible within some laboratories or publication systems, but it does not address the underlying coordination problem: technical capacities and institutional mechanisms for selection, verification, and reuse are evolving at different rates. A durable response may require more efficient review arrangements, differentiated repositories and certification procedures, and resources for examining high-value contributions. These institutional possibilities are developed in Section 10.

The existence of foregone possibilities does not imply that each unrealized idea has demonstrable value. Some proposed results will fail testing, duplicate existing work, or require revision before release. The opportunity cost relevant to this paper is therefore an uncertain and counterfactual quantity. Its importance is that withholding an output can have costs even when the resulting decrease in visible productivity improves a producer's immediate evaluative position. The potential conflict is between different forms and timescales of responsible knowledge production, not between diligence and an obligation to publish without restraint.

7.3 Disclosure and Verification Labor

AI-mediated production can make evidence of contribution more valuable while leaving the cost of producing that evidence unevenly distributed. A researcher who can explain which tasks were delegated to a model, which judgments remained human, and how a result was checked may make an artifact easier to evaluate. Contributor-role standards, provenance vocabularies, and authorship guidance provide relevant institutional resources for describing roles, histories, and accountability [18, 15, 13]. Their existence does not establish that exhaustive documentary disclosure is necessary or proportionate for every artifact. The practical question concerns the evidentiary work sufficient for a particular claim and assessment context.

Disclosure costs include the preparation of work histories, curating intermediate outputs, explaining rejected alternatives, retaining execution environments, communicating model limitations, and responding to questions from different audiences. Verification costs include tests of internal consistency, independent reproduction where feasible, review of sources and datasets, and additional assessment of safety or ethics where the domain warrants it. Some of this work is integral to responsible research independently of AI. Other parts become salient because observers cannot infer the production process from the final artifact alone. The distinction matters: the goal is a proportionate evidentiary basis for substantive judgment, with added requirements justified by the risks and purposes of the evaluation.

The same nominal standard may impose substantially different burdens on differently situated producers. A laboratory can allocate documentation and replication tasks across professional staff, collaborators, and shared computational infrastructure. A producer working without comparable organizational support may have to perform those functions within the same limited time and budget used for original research and basic maintenance. Consequently, an evidence requirement that appears identical at the level of formal rules can have asymmetric opportunity costs. Such asymmetries are not necessarily proof of improper institutional conduct; they are properties of the resources and division of labor available to participants.

A further difficulty concerns the object of proof. A complete log of model interactions is not equivalent to a proof of epistemic quality, and an unbroken version history does not by itself establish a researcher's ability to interpret the result. Extensive histories may contain confidential information, personal data, irrelevant exchanges, or materials that cannot be shared under applicable obligations. Evaluation can become less effective if reviewers receive large undifferentiated collections rather than targeted evidence of the relevant decisions and checks. The usefulness of a provenance record depends on its relationship to the evaluative question, its interpretability, and the independence of the available verification methods.

There is therefore a possible second-order burden of credibility production. To make intellectual output recognizable, the producer creates additional descriptions, demonstrations, audit materials, and explanations. These supplementary artifacts can themselves require examination, maintenance, and further explanation. Under insufficiently specified demands, the process may extend without a clear stopping rule. The tragedy of diligence is sharpened when the additional labor required to justify productivity grows alongside the productivity it was intended to make intelligible. An appropriate institutional standard should specify what question is being evaluated, which kinds of evidence can answer it, and when sufficient evidence has been provided.

The producer also faces a choice between broad transparency and selective disclosure. Broad disclosure may increase reconstructability in some cases, but it can expose intellectual property, third-party data, or preliminary errors without improving the decision at hand. Selective disclosure may protect legitimate interests and lower burdens while leaving an evaluator uncertain about essential production decisions. The scope of disclosure should accordingly depend on the risk profile, contribution claim, available alternatives, and safeguards for information that should remain restricted. The analysis supports intelligible and contestable evidence, with a proportionate burden of production, rather than a maximal demand for continuous documentation.

7.4 Visibility, Reputation, and Dependence

Producers do not enter evaluation systems through identical channels. Established institutions, professional affiliations, collaborators, disciplinary conventions, and previously formed relationships can affect which work is encountered and which explanations receive attention. These associations can carry legitimate information: a documented record of reliable review or successful collaboration may reduce some uncertainty for a potential evaluator. They can also operate as crude substitutes for directly examining a contribution when the producer's position is more visible than the evidence of the work itself. Signaling theory makes the use of observable indicators under uncertainty intelligible without establishing that a particular indicator is always accurate [21].

Institutional reputation and independent evidence must be distinguished. A recommendation based on the recommender's direct inspection of a method contributes something different from an affiliation displayed without reference to the work. Collaborative verification supplies a different basis for recognition from a formal title. If evaluators lack time to investigate these differences, they may rely on visible standing or prior familiarity to choose which portfolios warrant examination. Such selection can be efficient in some circumstances and exclusionary in others. Its adequacy depends on the relevance and reliability of the signal, the possibility of contesting a decision, and the availability of alternative routes to substantive assessment.

One response available to a relatively unfamiliar producer is to seek sustained interaction with communities capable of examining the work. Repeated technical discussion can make competence and responsibility observable in ways that a portfolio alone cannot. External feedback can also identify faults, improve methods, and create opportunities for joint production. Yet entering a network consumes time and may involve formal access requirements, travel, organizational fees, or expectations of unpaid participation. Connection itself cannot guarantee examination, trust, or material support. The value of an encounter depends on what information and reciprocal commitments actually become available through it. The more general mechanisms of heterogeneous encounters and power-conditioned observability require independent analysis; the present paper considers them only as possible interfaces with recognition.

Reputation-building can create additional dependence when evaluative pathways are concentrated in a small number of organizations or gatekeeping relationships. A producer may feel compelled to acquire a credential, accept an unfavorable allocation of work, or devote substantial time to visibility activities as a condition of being considered. Such outcomes are possibilities rather than assumptions about any specified institution. Their significance is that recognition may depend partly on access to social and material resources that differ from the underlying productive competence under evaluation. The connection between recognized contributions and sustainable resources is addressed more fully in Section 9.

Table 6 summarizes the distinct benefits, costs, and institutional dependencies associated with the producer responses considered in this section. It is an analytical typology, not a ranking of optimal strategies. Several strategies can be combined, and their consequences can change as evaluators, technologies, and productive relations evolve.

7.5 Feedback and Productive Sustainability

The preceding choice conflicts acquire particular significance when recognition influences the resources available for subsequent creation. A producer may need income, equipment, institutional access, maintenance time, or collaborators to continue producing and validating work. Recognition can facilitate some of these resources by making a contribution legible to an employer, a funding body, or a potential partner. Nevertheless, the relationship is contingent: a favorable technical evaluation does not create a funded position, and a useful public resource does not necessarily generate compensation for its maintainer. The feedback problem thus has at least two separable interfaces, one connecting output to recognized contribution and another connecting recognized contribution to sustaining resources.

A simple resource accounting relation can make the sustainability issue explicit without purporting to estimate behavior. Let F_t denote the resources available to a producer at the start of period t , expressed

Table 6: Potential benefits, burdens, and institutional dependencies of responses to uncertain recognition. The entries state conditional mechanisms rather than observed average effects.

Response	Potential benefit	Potential burden	Institutional condition
Accelerated release	More tested contributions become accessible	Evaluation congestion and continuing maintenance	Review capacity and provenance accessibility
Selective restraint	Focused examination and lower immediate load	Delayed inquiry and unrealized complementary applications	Criteria for selecting representative work
Process disclosure	More legible contribution and accountability	Documentation and confidentiality costs	Proportionate evidence standards
Independent verification	Stronger grounds for quality judgment	Time, compute, expertise, and coordination costs	Affordable and accessible review pathways
Relational engagement	Direct examination and iterative collaboration	Participation costs and possible dependence	Open access and substantive feedback opportunities

in a consistent resource unit; let Y_t denote resources acquired from paid work or other sustainable support during the period; let G_t denote externally supplied project support; and let E_t denote consumption and the net costs of production, communication, and verification. The period-end balance is represented by Equation 8. The symbols F_t and E_t are reserved here for the producer’s resource accounting, distinct from the evaluative backlog B_t and evaluation capacity C_t defined in Section 4.

$$F_{t+1} = F_t + Y_t + G_t - E_t. \quad (8)$$

Equation 8 is an accounting identity under its definitions, not a model of how recognition causes income. To preserve dimensional consistency, all terms must use the same unit and accounting horizon, with nonmonetary constraints such as available hours treated separately. If an analysis includes in-kind assistance, its valuation requires an explicit rule; time and money cannot simply be added together. The account nonetheless shows why a high rate of intellectual creation may coexist with declining productive reserves if material inflows remain insufficient relative to continuing costs. It also shows why an improved rate of recognition need not independently resolve the problem when funding eligibility, organizational budgets, and other conditions limit returns.

A second constraint concerns the use of finite attention by the producer. Let H_t represent the hours available for productive and supporting work, divided among creation h_t^c , validation h_t^v , communication h_t^m , and relationship or institutional engagement h_t^r . Equation 9 records the feasibility condition, with h_t^o representing other obligatory tasks within the chosen accounting window.

$$h_t^c + h_t^v + h_t^m + h_t^r + h_t^o \leq H_t. \quad (9)$$

The categories in Equation 9 are analytical conveniences: in practice an activity may serve multiple purposes, and consistent allocation rules would be needed for measurement. The constraint indicates a potential substitution problem. A producer who devotes more hours to explaining or verifying existing work may have fewer hours for new inquiries. Another who focuses on new outputs may have less time for interactions through which those outputs can be recognized. AI may change the rate at which tasks are performed but does not, by itself, eliminate every competition over human attention and access to other people.

These constraints can support several conditional feedback configurations. In a favorable configuration, efficient production is matched by proportionate evaluation, reliable adoption, and adequate resources for maintenance and further inquiry. Improved productive capacity can then reinforce sustainability. In an unfavorable configuration, additional production creates evaluative and maintenance demands while recognition and resources fail to expand sufficiently. Available time and reserves decline, which may weaken subsequent verification, communication, or participation. This can intensify the very

interpretive difficulties that limited recognition in the first place. Neither feedback direction follows from Equation 8 or Equation 9 alone; the relevant behavioral relationships must be specified and tested.

The term *tragedy* is justified here by a restricted feasibility argument. Given a finite horizon, binding limits on time or resources, and choices that each carry positive opportunity costs, an individual may lack a costless option that simultaneously maximizes production, validation, public dissemination, and relational engagement. Such a trade-off does not establish that all feasible options are harmful or that an institution could never expand the frontier through shared review infrastructure, income support, or better evidence organization. A stronger claim of unavoidable loss would require specifying objective functions, constraints, feasible alternatives, and the applicable welfare criterion. The present argument identifies a potentially burdensome allocation problem rather than proving a universal impossibility theorem.

Hypothesis 7.1 (Conditional productive-diligence dilemma). For producers facing binding material or time constraints, limited access to substantive evaluation, and positive costs of verification and institutional participation, increasing productive throughput may enlarge the set of feasible intellectual contributions while simultaneously increasing the costs of securing reliable recognition. Selective restraint may reduce some immediate evaluative burdens while increasing the opportunity cost of deferred contributions. The strength and direction of this conflict depend on institutional assessment practices, the evidentiary organization of portfolios, and available supporting resources.

Hypothesis 7.1 deliberately states a conditional conflict rather than a theorem of universal disadvantage. It also differs from Hypothesis 6.1. Recognition inversion concerns a specific adverse change in evaluation as output rises, holding relevant capability and quality constant. The productive-diligence dilemma concerns an allocation problem that can persist even when evaluators respond fairly to available evidence. For example, a researcher may face substantial documentation and maintenance costs despite receiving a favorable review. Conversely, a genuinely poor result may justifiably receive negative evaluation without supporting either hypothesis. Conflating these outcomes would prevent the theory from identifying its mechanisms.

The empirical inquiry must therefore measure more than visible productivity and expressed trust. Relevant observations include effective working time, financial reserves, evaluation opportunities, time spent on disclosure and verification, the share of outputs substantively checked, actual quality indicators, and the timing of material support. Longitudinal designs could examine whether changing evaluation procedures alter these relationships, while controlled portfolio studies could separate the effect of output volume from verified quality and production-process information. The research designs require careful attention to reverse causation: limited resources may cause both high reliance on AI and difficulty obtaining institutional recognition. Section 11 develops these methodological issues.

The productive-diligence dilemma concerns the opportunity costs created when generation, verification, attention, and material support operate under different constraints. Increased output can require additional selection and maintenance; restraint can postpone useful results; documentation and professional visibility can consume resources needed for further inquiry. These costs are conditional and institutionally variable. Section 8 formalizes the narrower relation among output, examination, competence attribution, and recognition before Section 9 addresses the material-support interface.

8 A Parsimonious Model of Production and Recognition

The preceding mechanisms can be compared more precisely within a restricted formal model. This section assigns distinct roles to output quantity, artifact quality, evidence of contribution, examination probability, inferred competence, and institutional recognition. The production law and review-selection mechanism are specified before an illustrative attribution rule is introduced. Conditional results and limiting cases identify which effects follow from stipulated functional forms and which require observations of actual evaluators. In particular, an attribution penalty is a modeling possibility rather than evidence that AI-mediated production generally causes one.

8.1 Actors, States, and Assumptions

Consider a producer p , an evaluator o , and an institution j that defines a task of evaluation during an interval of specified duration. The producer may create, test, and release intellectual artifacts with a combination of human effort and computational assistance. The evaluator can observe a portfolio summary and selected pieces of evidence, but initially lacks direct access to the producer’s complete knowledge, judgment, and production history. The institution supplies the standards, available evaluation resources, and procedures that connect particular judgments to endorsements or opportunities. The roles need not correspond to distinct persons or organizations in every real setting. They identify functions that are analytically separable even when performed by the same participants.

The producer’s underlying task-relevant competence is represented by c , an unobserved attribute relative to a specified class of tasks. The symbol does not denote a context-free measure of intelligence or personal worth. Let h be human work time, a a vector describing the availability and mode of AI assistance, and v the human and computational resources allocated to validation. The submitted portfolio contains n artifacts, each with quality attribute q_i relative to an explicitly defined assessment standard, and possibly documentary evidence d_i about its production, testing, or provenance. The values q_i may be multidimensional. Where a scalar enters the illustration below, it stands for a task-specific summary and should not be interpreted as a universal measure of epistemic value.

An evaluator sees an information set D_o , possibly containing submitted artifacts, declared AI involvement, test results, selected work histories, and institutional signals. Recognition has several possible outcomes: the probability that a focal artifact receives substantive examination, the evaluator’s inferred competence, a favorable assessment conditional on examination, and an institutional decision. These outcomes cannot be collapsed into a single quantity without losing the distinctions established in Section 6. In particular, a lower probability of per-artifact examination does not itself imply a lower judgment of the producer’s competence.

Assumption 8.1 (Restricted comparison protocol). A comparison between a lower-output and a higher-output presentation holds constant the producer’s task-relevant competence and the independently validated quality of a designated focal artifact. It specifies the evaluator’s decision task, the presentation and ordering rules, the surrounding workload, and the evidence permitted to vary with the portfolio. The protocol records separately any change in review costs or additional evidence caused by the number of artifacts.

Assumption 8.1 defines a controlled theoretical comparison, not an assertion that actual producers can expand output without changing quality, effort, or documentation. It is necessary because two common empirical comparisons are different: one examines the same focal artifact in portfolios of different reported sizes, while another measures the aggregate returns from a growing portfolio of distinct artifacts. The model addresses both but keeps their denominators explicit. It also distinguishes a producer’s actual AI-supported production process from an evaluator’s beliefs about that process. A belief may be appropriately responsive to evidence, inadequately calibrated, or influenced by an irrelevant signal; the mathematical direction of an update does not by itself identify which description is correct.

8.2 Production and Artifact Distribution

AI assistance changes the feasible mappings from time, judgment, and computational resources to candidate artifacts and validated contributions. A minimal stochastic representation allows the number and qualities of artifacts to respond differently to those inputs. Equation 10 defines the joint conditional production law.

$$(N, \mathbf{Q}, \mathbf{D}) \sim P_{\text{prod}}(\cdot \mid c, h, a, v, \mathcal{I}), \quad (10)$$

where N is the count of artifacts satisfying a stated inclusion rule, $\mathbf{Q}$ their quality attributes, $\mathbf{D}$ their supporting records, and $\mathcal{I}$ the relevant research and institutional environment. Equation 10 is a notation for a family of possible production processes. It permits an intervention on a to increase expected N

while leaving some dimensions of $\mathbf{Q}$ unchanged, increasing others, or reducing still others. Its substantive content must be supplied by task-specific constraints, validation procedures, and data.

For example, the ability to produce additional candidate hypotheses may expand sharply while the capacity to collect observations required for testing changes very little. Another workflow may use computational assistance to perform additional verification, raising quality or documentary legibility without materially changing submitted counts. These are different directions within the joint distribution in Equation 10. Accordingly, an increased value of n should not be interpreted as evidence of either superior or inferior artifact quality without additional assumptions or observations. Research on writing assistance and AI-enabled research demonstrates the relevance of task differentiation, although the findings from those settings do not identify the parameters of the present model [19, 25].

For the controlled portfolio comparison, designate an artifact i^* with verified quality q^* and evidence d^* . The surrounding portfolio changes from n_1 to n_2 items while q^* remains fixed. Additional items can be heterogeneous, and the model does not presume that each is valid or useful. This design is valuable because it isolates two effects that often coincide in an uncontrolled setting: a change in the evidence available about a producer and a change in competition for scarce review attention. Where portfolio expansion changes the focal item's presentation, the design must model that channel explicitly rather than treating it as a change in competence.

8.3 Evaluation Selection and Review Congestion

The evaluator's first decision concerns which submitted artifacts receive substantive consideration. Section 4 introduced an aggregate-workload representation using incoming task requirements and available review capacity. Here a simpler, artifact-level selection law illustrates what finite attention can imply even without adverse beliefs about AI assistance. Let $\pi_i(D_o, \mathcal{J})$ denote the probability that artifact i is substantively examined under evidence D_o and the institution's selection procedure $\mathcal{J}$. If w_i is the review work required by item i and $K > 0$ is available review effort, the feasibility condition is given in Equation 11.

$$\sum_{i=1}^n w_i \pi_i(D_o, \mathcal{J}) \leq K, \quad 0 \leq \pi_i \leq 1. \quad (11)$$

Equation 11 constrains expected review effort in the specified interval. It does not require uniform review quality, random selection, or identical artifacts. If tasks require non-substitutable expertise, the scalar K would need to be replaced by a vector of resource-specific capacities. For the present argument, the scalar form is sufficient to establish that an increased number of demands can reduce some individual examination probabilities unless examination efficiency, time, or selectivity changes.

Consider a deliberately stylized case with exchangeable artifacts of equal review weight and k available examination slots. A background workload of $\ell \geq 0$ competing items is included. When the focal item is among n submitted portfolio items and selection is uniform, its examination probability can be represented by Equation 12.

$$\pi(n) = \min \left\{ 1, \frac{k}{\ell + n} \right\}, \quad n \geq 1, \quad k > 0. \quad (12)$$

The values k and ℓ are expressed as counts of equally demanding items under this illustration. On the congested branch $\ell + n > k$, the focal artifact's examination probability falls as the portfolio expands, provided the other conditions remain fixed. Such a decline is an *examination effect*, not an evaluative declaration that the producer is less competent. It would arise under uniform allocation even where every artifact was produced without AI and was independently verified. Hence any empirical attempt to identify an AI-specific credibility penalty must measure examination selection separately from judgments conditional on review.

Even this minimal example requires careful choice of outcome. A lower $\pi(n)$ can coexist with a larger expected total number of examinations received by a producer, since the aggregate expectation is

$n\pi(n)$. With $\ell > 0$ and $k < \ell + n$, that aggregate is $kn/(\ell + n)$, which increases toward k as n grows. Per-artifact exposure, total exposure, and exposure per hour of producer effort consequently have different comparative statics. A statement that a prolific producer receives *less recognition* cannot be justified merely by showing that each submitted artifact receives a smaller share of scarce review time.

8.4 Competence Attribution and Recognition

Competence attribution is an inference about a producer from partial evidence. A fully specified Bayesian approach would begin with a prior distribution over c , a likelihood for D_o conditional on competence and the production process, and a posterior $P(c | D_o)$. The present section does not assume that all evaluators perform such updating, nor does it assign a distributional form without data. The distinction between a warranted inference and a heuristic response to unusual productivity is instead examined through a reduced-form score with observable components. Such a score can represent an evaluator’s reported judgment under a controlled protocol, but it should not be equated automatically with actual competence or a calibrated posterior.

Let m_0 be baseline attributed competence under a specified evidence state, $\alpha \geq 0$ the conventional positive evidentiary weight assigned to visible production, and $\beta \geq 0$ the weight placed on a potentially adverse volume-related interpretation. Let $z \in [0, 1]$ index access to interpretable evidence about contributions, and let $u(z)$ be its possible direct contribution to attributed competence. Equation 13 defines an illustrative score for a portfolio of size n .

$$m(n, z) = m_0 + \alpha \ln(1 + n) - \beta(1 - z) [\ln(1 + n)]^2 + u(z). \quad (13)$$

The score is conditional on a specified production and evaluation environment. In a comparison across AI-use conditions, the coefficients α and β , the baseline m_0 , and the evidentiary function u may themselves depend on disclosed assistance, actual production arrangements, field conventions, and evaluator information. Equation 13 deliberately holds these contextual parameters fixed while varying portfolio size and legibility. Consequently, a negative volume effect derived below identifies a property of a stipulated evaluation rule; it does not, by itself, identify an AI-specific causal effect. Such identification requires additional comparisons across production and disclosure conditions, as discussed in Section 11. The quadratic term is a selected functional form, not a claim that doubt empirically rises as the square of log output. It allows a simple comparison between conventional positive productivity signals and a countervailing interpretation that becomes more influential in unusually large portfolios. The factor $1 - z$ depicts the hypothesis that accessible, interpretable contribution evidence can attenuate a volume-related discount. Neither the existence of that attenuation nor the effects of documentation are presumed universal. Provenance may expose genuine problems, evidence can be noisy, and the legitimacy of scrutiny depends on whether the relevant attribute is actually being assessed. A high z indicates potentially useful legibility, not guaranteed correctness or unlimited disclosure.

Institutional recognition requires an additional rule. Suppose an evaluator who examines the focal artifact provides a favorable competence-related judgment with probability $g(m)$, where g is an increasing, differentiable response function bounded between zero and one. For a transparent example, take $g(m) = 1/(1 + \exp[-\tau(m - \theta)])$ with $\tau > 0$ and threshold θ . The per-artifact probability of a favorable examined outcome is defined in Equation 14.

$$r(n, z) = \pi(n) g(m(n, z)), \quad g(m) = \frac{1}{1 + \exp[-\tau(m - \theta)]}. \quad (14)$$

The score $m(n, z)$ is interpreted as the component of the evaluator’s information that enters the favorable-judgment rule after the relevant examination stage. In an empirical setting, actual review may itself reveal information about quality or contribution that is associated with portfolio size. That informational change would require a conditional score $m(n, z, D_o^{\text{review}})$ and an explicit selection model. The present reduced form sets aside this feedback for tractability, so it cannot justify equating the outcomes observed among

examined submissions with those that would have been obtained had every assigned submission been examined.

Equation 14 factorizes examination and a favorable judgment conditional on examination; it is a probability product under those conditional definitions, not an assertion that examination and judgment are causally independent. The function g is a stylized decision rule. It is not a model of merit, validity, employment, or resource return. A recognized artifact may not yield material support, and an accurate assessment of an artifact need not settle how much of it should be attributed to a particular contributor. Those later transitions are examined in Section 9.

8.5 Conditions for Nonmonotonicity

The model permits two formally distinct adverse responses to expanding output. The first arises from changes in competence attribution m ; the second arises from the joint probability r of examination and favorable assessment. A sufficient condition for an attribution inversion follows by differentiating a continuous extension of Equation 13 with respect to n , holding z fixed. The actual count n is integer-valued; derivatives provide local signs in the continuous relaxation, while finite differences determine comparisons between feasible portfolios. Equation 15 records that derivative.

$$\frac{\partial m(n, z)}{\partial n} = \frac{\alpha - 2\beta(1 - z) \ln(1 + n)}{1 + n}. \quad (15)$$

For $\beta(1 - z) > 0$, Equation 15 is negative when $2\beta(1 - z) \ln(1 + n) > \alpha$. This is a *conditional attribution inversion*: the hypothetical evaluator's inferred competence declines with further reported output despite the modeled producer's fixed underlying competence and unchanged verified quality of the focal artifact. The derivation establishes a property of the score in Equation 13; it does not establish the psychological prevalence or normative legitimacy of the proposed response.

The illustrative score is also vulnerable to a form of explanatory circularity if its coefficient choices are interpreted carelessly. The negative quadratic component was selected precisely to permit a sufficiently strong adverse output response. Proposition 8.2 identifies the region in which that stipulated component dominates the favorable component; it is an algebraic threshold, not independent evidence that such a component exists in evaluators' actual judgments. The theoretical contribution lies in distinguishing observable pathways and specifying counterfactual tests, rather than in treating the existence of a declining region as a discovered behavioral regularity.

Proposition 8.2 (Attribution inversion in the illustrative score). *Under Equation 13, with fixed z , $\alpha \geq 0$, and $\beta(1 - z) > 0$, attributed competence decreases as n increases on the range where $\ln(1 + n) > \alpha/[2\beta(1 - z)]$. If $\beta(1 - z) = 0$, attributed competence is nondecreasing in n within this score specification.*

Proposition 8.2 follows directly from the derivative in Equation 15. For an entirely hypothetical parameterization with $\alpha = 1$, $\beta = 1/4$, and $z = 0$, the continuous score begins to decrease above $n = \exp(2) - 1$; with $z = 1/2$, the threshold moves to $n = \exp(4) - 1$. The numbers illustrate sensitivity to an evidence-access assumption and do not represent estimated thresholds in any scholarly population. The illustrative threshold shows how the adverse interpretation must be sufficiently strong relative to the positive conventional output signal for a decline in attributed competence to occur. A neutral or favorable response is equally possible under other parameters. It would also be possible to obtain a declining score for defensible reasons if unusually high output correlated with genuine noncompliance or weaker validation in the evaluator's task environment. Testing *unwarranted* attribution discounts therefore requires independent evidence about the artifacts, contributor tasks, and the calibration of the evaluator's information, not merely the observation of a negative score slope.

Recognition through the examination route has a different condition. On the congested branch of Equation 12, differentiate the logarithm of Equation 14. Equation 16 gives the resulting comparison.

$$\frac{\partial \ln r(n, z)}{\partial n} = -\frac{1}{\ell + n} + \tau[1 - g(m(n, z))] \frac{\partial m(n, z)}{\partial n}. \quad (16)$$

Table 7: Mechanisms and discriminating observations in the parsimonious production–recognition model.

Mechanism	Conditional prediction	Evidence needed for identification
Productive expansion	More candidate artifacts at fixed human time may be feasible	Generated, submitted, and validated counts; task and resource changes
Review congestion	Focal examination probability may decline as demands rise	Reviewer workload, selection rule, examination and delay records
Attributional response	Inferred competence may rise, remain stable, or decline with reported volume	Controlled quality, evaluator judgments, disclosure and provenance
Per-artifact recognition	Examination probability and favorable judgment may move differently	Separate attention and conditional decision measures
Aggregate recognition	Total favorable outcomes may rise despite falling per-item rates	Portfolio denominator, sampled items, institution-level capacity
Material return	Recognition may fail to generate sustainable producer resources	Adoption, eligibility, budgets, paid work and maintenance support

Equation 16 is negative whenever the second term is smaller than $1/(\ell + n)$. Consequently, per-artifact favorable recognition can fall even while attributed competence is rising: the positive effect of additional evidence may be insufficient to offset the decline in examination probability. Conversely, a sufficiently large rise in m can offset the congestion term. The first case does not constitute an attribution inversion. It is an examination bottleneck whose institutional significance depends on the evaluator’s objectives and the available alternatives.

An aggregate recognition measure introduces another distinction. If all focal-type items are exchangeable and share the same conditional favorable-assessment rule, their expected total number of favorable examined outcomes is $R_{\Sigma}(n, z) = n r(n, z)$. On the congested branch, the logarithmic derivative is given in Equation 17.

$$\frac{\partial \ln R_{\Sigma}(n, z)}{\partial n} = \frac{\ell}{n(\ell + n)} + \tau [1 - g(m(n, z))] \frac{\partial m(n, z)}{\partial n}. \quad (17)$$

The positive first term in Equation 17 reflects increasing aggregate examination opportunities up to a finite capacity limit when the background workload ℓ is positive. For $\ell = 0$, aggregate examination saturates at k under full congestion, and only the change in conditional favorable judgment alters the expected number of favorable outcomes. Thus a falling per-item rate need not produce a falling aggregate return. A decline in aggregate favorable outcomes requires a sufficiently negative attribution response or some additional mechanism omitted from the illustration. This is a stricter form of recognition inversion than the per-artifact definition.

8.6 Comparative Statics and Sensitivity

Table 7 summarizes the model’s distinct channels. It is intended to prevent an adverse observation in one dimension from being misreported as deterioration across all dimensions. In particular, review congestion, attribution discounts, and material-resource constraints are not interchangeable explanations, even when they occur within one institution.

Several limiting cases clarify the model. With ample examination capacity, $\pi(n) = 1$, and only the conditional assessment channel governs r . With a zero volume-related discount, $\beta = 0$, the competence score is nondecreasing; nevertheless, scarcity can lower focal examination probabilities. With $z = 1$ under Equation 13, the selected volume-discount term disappears, but this is a property of the illustrative functional form, not a guarantee that perfect provenance dissolves all uncertainty. Increasing k mitigates examination scarcity; it does not independently improve the accuracy of m . A higher judgment threshold

θ changes observed endorsement frequency without changing the underlying producer's competence. These cases demonstrate why a credible model of recognition must report institutional decision rules alongside the producer's artifact output.

The role of legibility deserves a further qualification. In Equation 13, increasing z weakens the selected adverse attribution component. In an actual setting, generating and checking contribution records can also consume producer and evaluator resources. If greater documentation increases the item-specific review requirement w_i in Equation 11, a program to improve attribution may increase congestion unless it also improves review efficiency or evidence organization. An intervention can therefore strengthen the conditional basis for trust while initially reducing the number of cases an institution can examine. This is a possible trade-off, not a necessary effect: standardized, selectively accessible records may instead make evaluation more efficient. The distinction supports proportionate disclosure requirements and motivates separate tests of information quality, documentation cost, and selection probability.

Potential estimation strategies must address selection and endogeneity. Observed n depends on competence, opportunity, field norms, access to assistance, and submission choices; observed judgments depend on which portfolios were examined and what the evaluators knew. A correlation between productivity and inferred competence thus mixes both production and selection processes. To isolate the attribution route, a controlled experiment could hold the focal artifact and its independent quality assessment constant while varying reported portfolio size and documented contribution evidence. To isolate the congestion route, it could vary the evaluator's review load while holding the evaluator's access to artifact-level evidence and assessment criteria fixed. Measurements should record whether a respondent actually examines a work, their competence inference, their quality assessment, and their willingness to pursue cooperation as separate outcomes. Section 11 develops these designs in detail.

There are substantive limits to the present formalization. The institution's review standard, the evaluator's information, and the producer's contribution structure can evolve together. Field-specific expertise may create disconnected review bottlenecks rather than one scalar capacity constraint. Evaluators may learn to interpret AI-mediated portfolios better over repeated interactions, while producers may strategically change disclosure, selection, or validation practices. Institutional affiliation and relational trust may condition selection independently of the signal score. These mechanisms can be added when a particular study requires them; their absence from the parsimonious model marks analytical boundaries, not claims of unimportance. The neighboring papers on heterogeneous trajectories, observer cognition, power structure, and relational circulation develop complementary questions without being prerequisites for the present section.

The model therefore keeps examination probability, competence attribution, and institutional recognition separate. Review congestion can lower the first without reducing the second; an adverse inference requires additional assumptions about how productivity enters the attribution rule. Neither result determines whether an institution will provide wages, grants, infrastructure, or continuing support. Section 9 investigates that further decision boundary.

9 Recognition, Value Realization, and Resource Return

The formal distinctions in Section 8 concern whether an artifact is examined and how its producer is evaluated. This section follows a further institutional step: the possible translation of validated contributions and favorable judgments into sustained access to time, income, facilities, and maintenance resources. It distinguishes credibility, endorsement, and adoption; identifies employment, grants, contracts, and public or commons-based provision as resource interfaces; and examines budgetary and eligibility constraints and unequal costs of participation. These interfaces explain why recognition can be a prerequisite for some forms of support without guaranteeing their provision. A general account of relational currency remains outside the present inquiry.

Recognition and resource return are linked contingently. A finding may be sound and well understood yet irrelevant to the immediate work of a particular organization. A useful open resource may be deployed by many parties without supporting the people who created or maintain it. Conversely, a person may be

remunerated for technically useful labor before receiving formal public recognition of any individual artifact. Treating these cases as distinct is necessary for understanding why improvements in productive capacity or epistemic credibility need not secure continued production. It also prevents the explanatory model of Section 8 from implicitly assuming that a positive assessment determines a funding decision.

9.1 Credibility, Endorsement, and Adoption

The movement from intellectual contribution to sustained production involves several evaluative and organizational decisions. At a first stage, a contribution may become accessible and receive substantive examination. An examiner can then form a claim-specific judgment about correctness, novelty, utility, or methodological adequacy. A positive technical judgment may improve the credibility of the artifact or the producer for a defined task. The judgment may also acquire an institutional form, such as publication, software certification, an accepted technical report, a documented replication, or a selection decision. Each transition has its own threshold and informational requirements. Recognition is accordingly an event or relation with a specified object and audience, rather than an undifferentiated stock of prestige.

Substantive recognition provides information that a prospective user or employer may find relevant. Its consequences depend, however, on the decision for which that information is consulted. An evaluator can judge a mathematical method valid without requiring it in the evaluator's own research. An organization may recognize that a software implementation performs correctly yet lack the staff or systems needed to integrate it. A funding body may acknowledge the value of a program of inquiry while concluding that its announced call concerns another topic. None of these outcomes entails that the initial evaluation was insincere or technically defective. The relevant institutional action requires further premises regarding compatibility, obligations, opportunity costs, and available resources.

Adoption itself has multiple meanings. An artifact can be read or downloaded without being used. It can be evaluated experimentally without becoming part of a continuing workflow. It can be incorporated into a larger system while the initial developer remains uninvolved. It can also generate reciprocal relationships in which implementers, maintainers, and users exchange corrections, domain knowledge, resources, or commitments. These distinctions are especially consequential for research software, where initial implementation is only one component of long-term reliability. Studies of research software sustainability emphasize discoverability, usability, adaptation, maintenance, and institutional support in addition to availability [2, 14].

The present analysis therefore distinguishes at least four outcomes: an epistemically favorable assessment of a contribution; an operative institutional endorsement; use or adoption within a specified practice; and a commitment of resources to a producer, team, or maintained artifact. An increase in the first outcome may facilitate the others but need not do so, and the ordering can vary. For instance, practical adoption can lead to subsequent formal evaluation, while a grant can finance the evaluation and development of a proposed contribution before it has produced an independently validated artifact. The conditional ordering matters: a model that treats resources as the automatic terminal consequence of recognition would fail to represent these familiar cases.

A further distinction concerns the recipient of recognition and the recipient of resources. An organization may benefit from an adopted research tool even if the original creator receives neither compensation nor access to future decisions. A grant may support an institution's research program, with funds restricted to eligible personnel, equipment, or project activities. Credit can attach to the authors of a report while maintenance is carried out by other contributors. It is consequently insufficient to ask whether an artifact creates or realizes value in the aggregate; an inquiry into productive sustainability must specify *whose* conditions of continued creation are affected. This question is separable from judgments about whether a given distribution is ethically justified, although the distinction makes such judgments possible.

These interfaces have historical and institutional specificity. The National Academies' report on open science observes tensions between the public benefits of sharing research products and the incentives through which individual researchers are recognized and rewarded [16]. Such tensions arise in different

Table 8: Illustrative interfaces connecting recognized contributions to continued productive resources. The table distinguishes the object of institutional commitment from constraints on the producer’s actual access.

Interface	Basis for commitment	Possible supporting resource	Principal constraint
Employment or appointment	Fit with an authorized role and demonstrated capacity	Predictable income and institutional access	Vacancy, eligibility, role requirements, budget
Project grant or fellowship	Eligible proposal under funder criteria	Project time, expenses, or stipend	Call scope, selection, expenditure restrictions
Paid collaboration or contract	Agreed deliverables and reciprocal obligations	Fees, development time, maintenance support	Purchaser demand, procurement, enforceability
Research infrastructure provision	Expected shared or operational utility	Computing, facilities, staff, long-term stewardship	Governance, recurring operational budgets
Public or commons support	Agreed collective benefit and participation rules	Shared labor, grants, donations, pooled services	Coordination, stewardship, uneven contribution

forms across research articles, data, educational resources, and software. The report does not prove that a particular creator has been denied deserved compensation. It identifies an institutional problem relevant to the present model: outputs can be socially useful under arrangements whose career and funding criteria recognize a narrower subset of the work needed to create and maintain them.

9.2 Resource-Return Interfaces

Material return is mediated by arrangements with different purposes, time horizons, and decision criteria. Employment can finance continuing work through a wage or salary, usually in exchange for organizationally specified responsibilities. A research grant may underwrite a project with defined eligibility, expenditures, milestones, and reporting duties. A service contract can compensate deliverables or technical maintenance under an agreed scope. Institutional provision can supply equipment, secure data access, review labor, computing resources, or infrastructure support without transferring unrestricted income to an individual. Commons-based arrangements may combine shared maintenance, volunteer contributions, public budgets, memberships, and other forms of support. These mechanisms do not constitute interchangeable currencies. Their restrictions, beneficiaries, and renewal conditions differ substantially.

Table 8 presents representative interfaces and the distinct commitments they involve. The table is a conceptual classification of possibilities, and it does not rank them or presume their availability in any particular jurisdiction. A positive evaluation can sometimes be relevant to more than one interface, but an institution must still have authority and capacity to activate the corresponding arrangement.

The table makes a useful separation between recognition as evidence and recognition as an institutionally operative decision. A technical endorsement may serve as one input to hiring, but a favorable technical opinion is not an offer of employment. Conversely, an employer may acquire enough task-relevant information through a supervised work sample or collaborative trial to make an appointment without a history of conventional publications. A contract can support a maintainer’s work because an organization requires reliability, documentation, and repair services, independently of the number of new scientific results produced. Recognition of continuing maintenance work may therefore have materially different consequences from recognition of a one-time artifact.

The role of research software provides a particularly useful institutional comparison. An international survey of research funders reports varied mechanisms for supporting research software and describes attempts to improve recognition, career pathways, and maintenance funding [14]. These observations indicate that institutional responses can be developed beyond general endorsement of software’s importance. They also show why a single model of academic rewards would be incomplete. However, the

respondents constitute a particular sample of funders reporting their policies and perceptions; the survey does not establish a universal causal effect of a given support mechanism on developer livelihoods.

To express the conditional relation without presuming a single mechanism, consider a specified producer p , institutional interface j , and observation interval t . Let $C_{pj,t} = 1$ denote an operative commitment capable of transferring usable resources to the producer during that interval, and let $M_{pj,t} \geq 0$ denote the amount of such resources valued in a consistently defined accounting unit. The conditional expected material return is decomposed in Equation 18.

$$\mathbb{E}[M_{pj,t} \mid D_{pj,t}] = \Pr(C_{pj,t} = 1 \mid D_{pj,t})\mathbb{E}[M_{pj,t} \mid C_{pj,t} = 1, D_{pj,t}], \quad (18)$$

where $M_{pj,t}$ is defined to be zero absent such a commitment, and $D_{pj,t}$ includes relevant assessment evidence and institutional conditions. Equation 18 is an elementary conditional-expectation identity under these definitions. Its purpose is conceptual: evidence of competence may increase willingness to commit without guaranteeing that a usable commitment occurs or that its amount is adequate. The identity is neither an estimated funding law nor a presumption that money, in-kind support, and time can be combined without explicit valuation conventions.

An institution's decision process can be represented in greater detail by conditioning on demand, available budgets, eligibility, permissible uses, and decision authority. Yet adding a complex conversion calculus would obscure the immediate objective of this paper. For present purposes, the institutional interface specifies the point at which a favorable evaluation becomes an executable commitment that can help sustain production. General theories of how reputational claims, commitments, and relationship histories circulate across different interfaces belong to the companion investigation of relational currency.

9.3 Constraints Beyond Epistemic Trust

A high degree of warranted epistemic confidence is insufficient for material return where organizational demand, authorization, or resources are absent. This conclusion does not depend on the recognition-inversion hypothesis. Even in a setting where all evaluators accurately assess a researcher's competence and all artifacts receive adequate review, a producer may remain without stable income. A theory of recognition must therefore keep separate the reasons an evaluator believes a contribution to be reliable and the reasons an institution can assume a continuing material obligation. Epistemic adequacy is a property of a claim or judgment; budgetary commitment is an organizational action.

Demand constraints are one class of explanation. Useful research can have diffuse potential beneficiaries while no single organization has a present incentive or mandate to pay for its production or upkeep. Conversely, an organization may have a pressing problem for which even a modest, specialized contribution has immediate value. These differences are not reducible to a general hierarchy of scientific quality. Some contributions serve as enabling methods for other projects, and their value may be realized indirectly, after a delay, or by users outside the creator's organizational environment. The possibility of such delayed and distributed benefits complicates the inference from public use to individual compensation.

Eligibility and authority constraints form another class. Hiring and grant decisions are often subject to formal procedures, residency or disciplinary requirements, personnel classifications, security or ethics obligations, and restrictions on expenditure. A committee may judge an applicant's work favorably while lacking authority to create a position. A funder may support software infrastructure but prohibit its award from covering certain maintenance or personal expenses. Some restrictions protect legitimate public interests and accountability; others may be poorly matched to evolving methods of research. Their consequences must be assessed according to their actual legal and institutional purposes. Successful persuasion of an evaluator cannot by itself alter every condition governing resource allocation.

Rights and access conditions may further separate value realization from the producer's return. Open licenses can enable reuse with little direct compensation, while restrictive rights can limit adoption and downstream improvement. Legal entitlement to particular uses and the distribution of benefits vary by license, contract, jurisdiction, and institutional ownership. An open artifact may generate substantial

public benefits through repeated use or integration into complementary systems. Those public benefits do not automatically establish that a funding channel for the maintainer exists. At the same time, openness may attract collaborators, reputation, employment opportunities, and external support. The paper accordingly treats open availability as a condition that can alter the structure of feasible interactions, with its net return depending on the arrangements through which use and support are organized.

The institutional literature on open science articulates this tension without implying that openness is inherently financially disadvantageous. The National Academies report describes how incentives for promotion, credit, and the sharing of data and code can diverge [16]. UNESCO's *Recommendation on Open Science* likewise presents open knowledge, infrastructure, equitable participation, and supportive policy as related institutional concerns [24]. These are normative and policy materials rather than causal estimates of individual returns. Their relevance here lies in demonstrating that resource and governance arrangements must be considered alongside the accessibility of scientific outputs.

Recognition is also heterogeneous across audiences. A computational artifact may be highly regarded by an expert user while receiving limited weight in a recruitment procedure emphasizing different credentials. A study may be widely discussed while a local committee considers it outside an authorized program. This plurality does not warrant assuming that one judgment is necessarily illegitimate. It indicates the need to specify the criterion, audience, and consequence when referring to *institutional recognition*. Where standards are unnecessarily opaque or unrelated to the activity being evaluated, those features can become subjects of institutional critique and empirical research. A complete explanation requires evidence about the actual procedure rather than imputing motives from its outcome.

Finally, there is a temporal constraint. Recognition may arrive after the producer's capacity to maintain a project has declined, while long-duration obligations require support at regular intervals. A large but uncertain prospective grant can have less immediate sustaining effect than a modest reliable income stream when unavoidable expenses are due earlier. Nonfinancial opportunities may also require time or equipment the producer cannot supply. Consequently, an account of resource return should record the timing, certainty, restrictions, and beneficiary of support, not simply whether recognition eventually occurs.

9.4 Differential Costs of Participation

The interfaces described above are encountered under unequal productive conditions. A participant with secure income, dedicated staff, established collaborators, and access to specialized infrastructure can allocate additional time to verification, professional interaction, and funding applications without necessarily jeopardizing continuing research. A participant who relies on short-term earnings and individually financed tools may encounter substantially higher opportunity costs for the same activity. These are analytical contrasts between resource configurations. Their prevalence and distribution in any population require evidence beyond the hypothetical comparisons developed here.

The problem becomes visible when the nominal price of accessing an evaluation opportunity is low but its full cost is high. Preparing a detailed application may involve technical documentation, translations, travel, reference letters, access to computing resources, formal demonstrations, or repeated meetings. A procedure described as open can impose considerable time requirements and requirements for prior relationships. The burden of such conditions depends upon prior available resources. That dependence is particularly relevant to AI-mediated knowledge production because inexpensive generation of candidate artifacts does not imply inexpensive validation, documentation, or institutional participation. The producer may save time in one task yet face increasing requirements in another.

Provenance infrastructures can partly alleviate these burdens. Version-controlled code, reproducibility packages, structured contribution declarations, and preserved review records may make selected claims easier to examine. However, an exhaustive requirement to document each computational interaction could become costly or intrusive, especially where private data, collaborator confidentiality, or intellectual-property obligations are involved. The appropriate question is the level of evidence sufficient for a defined evaluative claim. The work of making an artifact auditable ought to be distinguished

from demands for complete reconstruction of a person’s intellectual history. Section 10 will assess proportionate evidence standards and shared evaluative infrastructures more directly.

The possibility of unequal evaluation burdens can be represented within the resource account of Equation 8. Let an opportunity have a positive net expected resource return over a specified horizon but require up-front expenses or time before the return can be realized. A producer with sufficient reserves may enter and bear the uncertainty; another with the same competence may be unable to cover the initial cost. This creates a selection effect even if eventual evaluations are accurate and procedures apply nominally identical quality criteria. The resulting differences in observable recognition would reflect both evidentiary judgments and unequal feasibility of participation. Identifying the selection effect empirically would require data on who was eligible, who applied, who could complete the application, and who received an offer, rather than information about successful applicants alone.

There are also continuing costs after adoption. A research resource may require security patches, documentation, user support, adaptation to changing dependencies, scientific validation, and accommodation of new use cases. These tasks can be valuable while producing few new countable publications. A producer evaluated principally by the quantity of visible new artifacts may face an unfavorable allocation problem: time spent preserving the utility of existing work can be less visible than time spent releasing a new result. Research software studies and policy discussions have identified maintenance, career recognition, and recurring support as practical challenges [2, 14]. The relevance is especially direct for scientific infrastructures whose value emerges through repeated downstream uses.

These considerations motivate a hypothesis that remains separate from recognition inversion. The latter concerns adverse changes in the evaluation of competence or output under controlled conditions. The present hypothesis concerns the distribution of effective participation and resource return after a contribution has already received favorable assessment.

Hypothesis 9.1 (Recognition–return decoupling under institutional constraints). When access to material support depends on additional organizational commitments, eligibility rules, budgets, or positive participation costs, a favorable epistemic evaluation of a contribution may fail to produce resources sufficient for the producer’s continuing participation. The likelihood and severity of this decoupling may vary with institutional design, the timing and restriction of resources, and the producer’s prior reserves, even where competence and verified artifact quality are comparable.

Hypothesis 9.1 does not imply that each recognized contribution creates an enforceable claim to payment, nor does it establish a single appropriate rule for distributing research resources. The hypothesis identifies variables through which epistemic and economic outcomes may diverge. A discriminating empirical design would measure verified quality, substantive evaluation, type of endorsement, actual adoption, resource-bearing commitments, net accessible support, and a producer’s capacity to continue work over a defined interval. Comparisons across institutions would also need to control for demand, field, project maturity, eligibility, and local costs. Strong evidence of recognition without material return could then be distinguished from insufficient attention, unfavorable technical judgment, and decisions motivated by legitimate resource scarcity.

Recognition and material return depend on related but distinct decisions. A positive technical assessment can coexist with restricted budgets, weak organizational demand, eligibility barriers, or continuing maintenance burdens. An account of sustainable intellectual production must therefore examine both the validity of evaluation and the availability of practical support. Section 10 assesses candidate institutional arrangements at these interfaces, leaving the circulation of relational credentials and value across social systems for separate research.

10 Institutional Adaptation and Evaluative Infrastructure

The preceding analysis identifies three institutional questions: which contributions receive examination, how evaluators infer responsibility and competence, and which decisions provide resources for

subsequent work. This section considers possible responses at each interface. It examines quality-sensitive evaluation, proportionate provenance, independent verification and contestation, substantive professional interaction, review capacity, and continuing material support. The proposed arrangements face tradeoffs among reliability, administrative cost, privacy, and equitable access. Their effectiveness is an empirical matter requiring attention to the purpose and resources of the institution implementing them.

Institutional adaptation is already under discussion and implementation in several research settings. The San Francisco Declaration on Research Assessment (DORA) advocates attention to the intrinsic merits and diverse contributions of research, with responsible rather than mechanical use of quantitative indicators [6]. The Agreement on Reforming Research Assessment, associated with the Coalition for Advancing Research Assessment (CoARA), similarly proposes qualitative evaluation with peer review at its center and recognition of diverse research outputs and activities [4]. These are institutional reform commitments and guiding principles, not experimental demonstrations that particular assessment procedures eliminate the effects modeled here. Their relevance is that they provide an existing policy context in which the evidentiary meaning of AI-mediated output can be examined without claiming the need to replace all established evaluation institutions.

10.1 Evaluation of Validated Contributions

A first institutional choice concerns the object of an evaluative decision. A hiring committee, journal editor, research funder, or technical collaborator may seek evidence about different attributes: the validity of an artifact, a person's understanding of a method, the originality of a contribution, the reliability of an implementation, or the feasibility of future collaboration. An undifferentiated evaluation of productivity risks combining these attributes into a single impression. In AI-mediated production, that combination may be particularly fragile because the number of completed artifacts supplies weaker evidence about the amount and type of intellectual work contributed by any specific person. A defensible procedure begins by identifying the decision-relevant claim before specifying the evidence required to assess it.

Consider an institution evaluating a candidate who submits a large collection of research artifacts. The institution might require the candidate to select a small number of representative contributions, identify the technical claims for which responsibility is accepted, and supply evidence proportionate to those claims. Such evidence can include a reproducible computational result, a carefully reasoned argument, an intelligible method description, or a technical discussion in which the candidate explains limitations and resolves questions. The evaluation remains open to contributions made across a broader portfolio, but it directs scarce attention toward substantiated examples rather than equating portfolio size with achievement. The institution's prior specification of criteria reduces the risk that a change in output count alone silently changes the standard applied to an applicant.

The choice of representative artifacts creates a selection problem of its own. Applicants may choose unusually favorable examples, while evaluators may overlook systematic defects in work that was not selected. A proportionate design can combine applicant-chosen evidence with limited independent sampling, field-specific checks, and disclosure of material limitations. In a safety-sensitive setting, the relevant sampling and validation requirements may be much more extensive than in preliminary theoretical inquiry. Nor does the availability of a strong demonstration of competence establish the correctness of every artifact in a portfolio. Validity remains claim-specific. The practical objective is to make the inference from evaluated evidence to an institutional decision explicit and contestable.

This design is compatible with the distinction among competence, contribution, and artifact quality established in Section 2. A competent researcher may produce a flawed artifact, while a valid artifact may depend heavily on computational generation, external datasets, or specialized collaborators. A research team may be responsible for a valuable collective result without each contributor possessing identical competencies. Accordingly, an institution should specify when it evaluates collective accomplishments, individual responsibility, or demonstrated task performance. The CRediT taxonomy provides one existing vocabulary for describing differentiated contributor roles, although role attribution alone does not

prove correctness or determine authorship status [18]. Reform initiatives that value software, datasets, methodological development, and reproducibility work supply a broader context for recognizing diverse forms of contribution [6, 4].

The institutional difficulty is not removed by replacing quantity metrics with a single holistic judgment. Qualitative assessment can reflect substantive expertise, but it can also reproduce informal prestige hierarchies, unrecorded assumptions, and inconsistent standards. Deliberation should therefore be accompanied by accessible criteria, task-relevant evidence, documented reasons in consequential cases, and appropriately limited opportunities for correction. These measures may reduce arbitrary recognition discounts while preserving the evaluator’s obligation to reject insufficiently substantiated claims. Their effectiveness remains an empirical question, to be examined through the designs proposed in Section 11.

10.2 Proportionate Generative-Trajectory Legibility

A second design problem concerns the production history that can be made available to an evaluator. Section 5 established that the same finished artifact can be consistent with distinct human and computational contribution histories. When a decision depends on the origins of a method, the distribution of intellectual responsibility, or the reliability of a result, selected traces of the production process may improve evaluation. Possible records include dated research questions, design specifications, version histories, source and dataset identifiers, validation reports, known-failure registers, and declarations of the activities in which AI systems were used. Interoperable provenance models such as W3C PROV-O provide a means of representing entities, activities, agents, and derivation relations [15]. These relations improve inspectability when appropriately populated; the formalism itself does not certify epistemic quality.

The required depth of disclosure should be indexed to a defined evaluative purpose. For a mathematical claim, a checkable derivation and clear assumptions may be more important than a complete transcript of exploratory dialogue. For a research-software artifact, tests, dependency versions, validation datasets, and a record of material design decisions may be more useful than exhaustive reconstruction of every edit. For a collaborative contribution, role descriptions and documented decisions may help establish responsibility, provided they are not presented as proof of private thought or personal competence. The central aim is *purpose-relative legibility*: enough interpretable evidence to evaluate the particular claim for which recognition is sought.

An institutional requirement for exhaustive histories could produce a new form of evaluative congestion. Reviewing extensive logs takes time; preserving them can entail storage, privacy, and security costs; and explaining automated or iterative processes can demand specialist labor. A procedure that regards greater documentation volume as an automatic proxy for credibility could re-create the quantity-based distortions that it was designed to mitigate. Moreover, research involving confidential collaborators, sensitive records, proprietary access restrictions, or unresolved intellectual-property questions may be unable to disclose all relevant materials publicly. Graduated disclosure, secure examination, redacted records, and verifiable summaries can be considered where they protect legitimate interests while permitting meaningful scrutiny. The sufficiency of those mechanisms depends on the claims being checked and the threats they are intended to address.

The unequal costs of these requirements deserve explicit consideration. A large research organization may have dedicated personnel and software for maintaining traceable workflows. An independent contributor may perform the same recordkeeping, testing, and explanation alongside the substantive research itself. When compliance costs are treated as outside the evaluation system, formally identical rules may impose markedly different burdens. Institutions could mitigate these differences through standardized, lightweight evidence packages, transparent minimum requirements, shared verification facilities, accessible templates, and opportunities to provide alternative evidence of equivalent probative value. The costs and benefits of such measures require evaluation rather than presumption. A lower documentation threshold can improve accessibility while leaving difficult cases less fully evidenced; a higher threshold can increase checking power while excluding participants unable to comply.

Trajectory legibility should also be distinguished from continuous personal surveillance. A contrib-

utor’s claim to recognition does not normally entail unrestricted access to private thought, communications, or every experimental failure. Knowledge of a production process remains partial and context-dependent even under extensive logging. A proportionate framework acknowledges this limit by making the requested inference and supporting record explicit. The wider theory of how an observer interprets a developing relational trajectory, including selective attention, retrospective reinterpretation, and cognitive bias, belongs to a separate investigation. The contribution here is restricted to institutional conditions under which selected evidence can become available for responsible assessment.

10.3 Independent Verification and Contestable Recognition

Verification provides a third form of institutional adaptation. AI-mediated generation can enlarge the number of candidate claims without ensuring a corresponding increase in checks. Evaluation arrangements may consequently benefit from allocating appropriate resources to independent testing, methodological replication, error analysis, and explicit reporting of uncertainty. The distinctions between computational reproducibility, the obtaining of consistent results in new studies, and broader evidentiary judgment are important here; they should not be collapsed into a single label of “verified” [17]. An internally reproducible computation may retain invalid assumptions, while the inability to reproduce a result may reflect missing information, environmental variation, or an actual defect. The outcome of any check should identify its domain and limitations.

A defensible verification arrangement links confidence to the evidence actually produced. In a computational example, executable tests can establish behavior under specified inputs and environments, yet they may leave the suitability of the scientific model unexamined. A formal proof checker can validate a formal derivation relative to selected axioms and definitions; it does not automatically settle whether those assumptions adequately represent the intended phenomenon. Independent domain expertise may be needed to examine experimental design, construct validity, or the relation of an artifact to prior work. Institutions should therefore record the object, extent, and result of each check, and distinguish automated checks from expert interpretation. This approach could reduce both unwarranted dismissal of AI-mediated work and unwarranted endorsement based on polished presentation.

Contestability addresses a separate institutional problem: evaluators can err, adopt inadequate inference rules, or receive incomplete evidence. In consequential decisions, a proportionate challenge procedure can permit the affected contributor to clarify a claim, submit corrected evidence, identify a mistaken attribution, or request review of a material procedural error. Such procedures need not require that an institution reconsider every scientific judgment indefinitely. They can specify grounds for reconsideration, deadlines, conflict-of-interest safeguards, the authority of decision makers, and the possible outcomes of a challenge. An organization may also retain the ability to withdraw or revise earlier recognition when new evidence discloses serious problems. The key issue is how justified judgments become revisable under changing evidence.

The relationship between independent verification and contestability can be illustrated by three distinct cases. In one, a newly discovered error justifies reducing confidence in an artifact. In a second, an evaluator discounts a correct result solely because its drafting involved AI, despite independent evidence adequate for the decision being made. In a third, a claim may remain unverified because appropriate expertise or data are unavailable. The first calls for correction, the second for scrutiny of the evaluation rule, and the third for calibrated suspension of judgment. An institution that treats these cases as equivalent would obscure the difference between evidence of failure and the absence of sufficient evidence. A reliable procedure should preserve these distinctions in its records and decisions.

Verification infrastructure also has a distributional dimension. Technical tests, secure data access, independent replication, and expert review have costs that must be borne by some actor. Placing those costs exclusively on individual producers may discourage valuable, high-risk, or public-interest work. Placing all costs on evaluators may be infeasible under large volumes of submissions. Shared test environments, reusable validation suites, staged review, and publicly documented standards are potential responses, subject to confidentiality and resource constraints. These proposals are consistent with the

wider emphasis on transparency and responsible research assessment [4], but their causal effects on recognition accuracy and participation inequalities remain to be established.

10.4 Relational Encounter and Substantive Assessment

A fourth possibility concerns the context in which a contributor becomes an object of evaluation. Portfolios and written applications present selected traces of intellectual production, but they cannot always disclose the contributor's ability to clarify a difficult problem, respond to criticism, revise a model, or cooperate in an unfamiliar technical setting. Structured discussion, live explanation, joint review of a research artifact, and short collaborative exercises can provide additional evidence. These encounters are especially relevant when the institutional decision concerns an ongoing relationship, such as employment or research collaboration, rather than the correctness of an isolated proposition.

A hypothetical comparison illustrates the limited claim. Two applicants supply portfolios of comparable documented quality. An evaluator who sees only the portfolios may assign similar judgments or rely upon institutional affiliations. During a structured technical discussion, each applicant may demonstrate different strengths in explanation, criticism, error correction, or collaborative planning. The resulting decision could be better informed if these attributes are genuinely relevant and if the interaction is evaluated under transparent criteria. The exercise can also mislead if performance is strongly affected by language proficiency, anxiety, cultural expectations, inaccessible format, or prior access to preparation and mentoring. Relational contact therefore supplies additional observations without inherently producing more valid evaluation.

Encounters may also introduce evaluators and contexts that were absent from an existing relationship network. A participant's prior trajectory can become legible to a new observer because a professional exchange creates an opportunity to inspect it. The same event can activate different interpretive categories or invite new forms of cooperation. These are possibilities for altering the conditions of recognition, not automatic trust increments. Institutional design can increase the accessibility of such encounters through open seminars, structured research discussions, cross-disciplinary review, and clearly advertised application routes. Opportunities mediated exclusively by personal endorsements or inaccessible social spaces may instead reinforce prestige advantages and selective visibility.

The present paper addresses relational encounters as a restricted institutional response to recognition difficulties. It does not develop a general probability law connecting event heterogeneity, informational surprise, attention, and trust. Nor does it model how observer-specific cognitive biases or broader power structures govern the interpretation of relational trajectories. Those are distinct research objects of the accompanying trajectory, cognition, and power investigations. Here the relevant decision criterion is comparatively narrow: whether an interaction supplies additional, task-relevant, fairly obtained evidence or facilitates access to evaluators who can competently examine existing work.

10.5 Review Capacity and AI-Assisted Evaluation

A fifth adaptation concerns the organization of evaluative work itself. In the constrained-capacity account of Equation 11, opportunities for substantive examination depend on available reviewer effort, task difficulty, and selection procedures. Institutional capacity can be expanded through additional staff, shared specialist resources, standardized checks, or processes that reduce avoidable clerical work. AI systems may help with particular limited tasks, such as checking internal document consistency, assisting with code testing, detecting duplicated passages for human inspection, or organizing publicly available information. Such tools can be useful without serving as independent arbiters of scientific novelty, authorship, or individual competence.

The difference between technical assistance and delegated judgment matters because evaluation involves confidentiality, contested standards, and accountability. Editorial guidance already differentiates acceptable assistance from uses that imperil the integrity or confidentiality of review; arrangements vary among publishers and settings [22]. For example, a reviewer may be prohibited from uploading an unpublished manuscript to an external AI service, even where that service might perform a preliminary

summary. In other settings, an institution may approve a controlled environment for limited internal checks. Evaluation design should specify data permissions, source reliability, error rates, security, appeal mechanisms, and the human official responsible for decisions. A general commitment to “AI-assisted review” does not resolve these details.

Automating early screening can increase the number of items handled while also concentrating errors. An opaque classifier that uses publication volume or institutional affiliation as a proxy for competence could reproduce the recognition inversion identified in Section 6. Likewise, a system trained chiefly on previously recognized contributors may perform poorly on unfamiliar production styles or interdisciplinary artifacts. The relevant performance measures should therefore include false dismissals, false endorsements, differences across research contexts, and the ability to identify errors in the selection procedure. They should also distinguish the speed of preliminary routing from the correctness of final expert assessment. A faster pipeline can process more artifacts without improving evaluative accuracy.

The aim of differentiated review is to align scrutiny with stakes and evidence. A preliminary research proposal, a publicly shared exploratory proof sketch, and a clinically consequential decision-support system require substantially different forms of checking. Institutions can publish review categories that specify the assurance sought, acceptable evidence, confidentiality conditions, and thresholds for escalation. Such procedures may conserve scarce expertise for high-stakes or technically unresolved questions while providing intelligible reasons for initial screening outcomes. However, inappropriate triage can exclude novel work precisely because it falls outside existing classifications. Periodic audit, random examination of screened-out material where feasible, and routes to submit alternative evidence are possible safeguards.

An evaluative system also has to acknowledge the work performed by reviewers and maintainers. Substantive assessment of machine-assisted outputs can remain labor-intensive, and high submission volumes may shift costs onto unpaid or poorly supported reviewers. Capacity expansion will therefore be incomplete if it consists solely of additional screening tools while leaving verification labor and expert responsibility without adequate time or resources. The institutional obligations and budgetary implications are considered in the following subsection.

Candidate reforms should also be compared against resource-constrained alternatives, including representative-portfolio assessment, direct technical demonstrations, independent validation, and staged requests for provenance. A reform that increases documentary requirements can increase accuracy for examined cases while excluding producers unable to satisfy the entry requirements. Evaluations of reform should therefore report accuracy, false endorsement, false rejection, application costs, the distribution of examination opportunities, and downstream support separately. Without these comparisons, an apparently more rigorous process may merely shift verification work to already disadvantaged contributors.

10.6 Material Support and Institutional Responsibility

Evaluative reform has limits when recognition remains disconnected from the resources required for participation. Section 9 showed that validated contributions can fail to generate usable income or sustained infrastructure support. Institutions concerned with valuable public knowledge work may therefore need to consider resource arrangements alongside assessment rules. Such arrangements can include salaried research roles, paid software-maintenance positions, commissioned validation work, accessible seed funding, shared computing facilities, compensated review, and mechanisms for supporting continuing public infrastructure. These are alternatives with distinct eligibility conditions and governance obligations, not components of a universal remuneration entitlement.

For research software, the maintenance of a validated resource may create ongoing technical obligations whose significance is poorly captured by the number of new papers or releases. Existing studies and institutional reports identify maintenance support, career recognition, and organizational funding as practical challenges [2, 14]. Support mechanisms should specify whether funds are available to an individual creator, a team, or an institution, and whether they can be used for ordinary living costs, maintenance, equipment, verification, or only narrowly defined project expenses. An award that supplies computing

Table 9: Candidate institutional responses to distinct recognition and sustainability constraints. Proposed benefits and failure modes are conditional hypotheses rather than established effects.

Response	Targeted mechanism	Potential limitation or adverse effect
Claim-specific portfolio review	Unexamined output-count signals and ambiguous competence attribution	Favorable selection of examples; inconsistent qualitative judgment
Purpose-relative provenance	Limited contribution legibility and incomplete verification evidence	Privacy exposure; documentation costs; unequal compliance capacity
Independent checking and challenge	Erroneous endorsement, unsupported discounts, and irreversible judgments	High review cost; strategic challenges; unavailable expertise
Structured professional encounter	Narrow observer access and limited evidence of task-relevant interaction	Social-selection bias; language and format barriers
Differentiated assisted review	Screening congestion and repetitive technical checks	Automation bias; confidentiality failures; opaque exclusion
Continuing resource support	Separation of credible contribution from sustainable participation	Budget constraints; capture; restrictive eligibility

access may be materially valuable yet insufficient to sustain an otherwise unfunded researcher. Likewise, a broadly recognized public tool can generate widespread downstream benefits without providing its maintainers with a stable income.

Allocation rules involve difficult questions of justice and organizational purpose. Public-interest research may justify collective support even when commercial demand is limited, while other projects may properly compete for restricted resources based on field-specific priorities and evidence. Support for independent contributors can improve accessibility, but poorly specified schemes can expose institutions to unsustainable commitments or weaken accountability. A careful analysis must therefore consider available budgets, the risk of capture by established groups, transparency of selection criteria, accessibility across institutional positions, and the distribution of maintenance obligations. UNESCO's Recommendation on Open Science explicitly situates open science within questions of infrastructure and enabling policy [24]; it does not, by itself, determine a feasible allocation formula.

Table 9 synthesizes the proposed response mechanisms and their potential failure modes. It deliberately separates the mechanism a reform is intended to address from evidence that the reform has succeeded. That separation is important because more elaborate documentation, more interpersonal contact, and more review technology can each create additional costs or inequalities if deployed without adequate safeguards.

The appropriate combination of measures depends on the purpose of the institution and the object of assessment. A small research group may be able to conduct detailed interviews but lack independent validation infrastructure. A large scientific archive may support reproducibility checks yet have no authority to compensate every contributor. A funding organization can subsidize verification and maintenance but must administer finite budgets and respond to competing public goals. Accordingly, the proposed institutional responses form a repertoire of interventions, not a single package to be implemented identically across domains. Their comparative value depends on whether they improve substantive recognition accuracy, reduce unwarranted exclusion, preserve responsible scrutiny, and strengthen the conditions for continuing participation.

These arrangements create distinct, testable institutional propositions. Adjustments to triage may affect who receives examination; provenance and substantive encounters may affect what evaluators infer; funding and employment procedures determine whether endorsement yields support. Section 11 specifies comparisons that can assess their effects and costs without assuming that any single reform will succeed across disciplines.

11 Empirical Research Designs and Competing Explanations

The recognition paradox requires tests capable of separating attention, inference, and material-return mechanisms. This section specifies controlled portfolio comparisons, disaggregated measures of evaluation, observational designs, and potential falsifiers. Particular attention is given to the selection of artifacts and candidates into review: outcomes observed only after screening may confound access with conditional judgment. The proposed designs seek to distinguish adverse responses to production volume or AI disclosure from quality differences, changing applicant populations, institutional demand, and legitimate uncertainty about contribution. They are research protocols, not reports of experiments conducted for this paper.

The core identification difficulty is that observable production volume is jointly shaped by competence, access to tools, institutional resources, research opportunities, and the decision to disclose or submit work. Evaluator responses also depend on the quality and visibility of the artifacts sampled. Consequently, an association between a large portfolio and unfavorable recognition can arise from several incompatible processes: scarcity of reviewer time, a changing interpretation of authorship, weak or misaligned work, or selective exposure to unusually strong and weak portfolios. The designs below therefore separate changes in *what is submitted*, *what is examined*, *what is inferred*, and *what decision follows*. This separation follows the formal distinction between focal-item examination and conditional competence attribution in Section 8.

11.1 Controlled Portfolio Evaluation

A first research design would present evaluators with fictional but internally coherent portfolios of knowledge-production artifacts. Suitable materials include a research-software package with executable tests and documentation, a mathematical argument with a separately checked derivation, or a short technical paper with an independently reviewed methods appendix. All identities, institutional names, and claimed accomplishments would be synthetic. Independent domain specialists would first assess the substantive artifacts under a common rubric and without the experimental authorship or productivity descriptions. Because their own assessments may be noisy or field-dependent, the study should predefine acceptable equivalence margins and record inter-rater disagreements, technical test coverage, and residual uncertainty. Equal average ratings are insufficient to guarantee exchangeability of artifact quality, and an artifact's validity does not settle the competencies attributable to every participant in its production. Their assessments would serve as preliminary evidence for comparable technical quality, rather than as guarantees that the artifacts have equal merit in every possible respect.

The evaluation experiment would independently manipulate at least four profile attributes: the reported number of completed artifacts over a specified period; the disclosure of AI assistance and its stated scope; the availability of inspectable provenance or contribution records; and independently corroborated evidence of artifact quality. A fifth attribute could vary the declared evaluation workload, with its own review-budget manipulation if the study seeks to identify congestion. For example, participants might examine the same selected software artifact under profiles reporting three or thirty additional completed items, with the same independent test results and with explicitly specified contributions by the human researcher. A volume-only contrast must keep the focal artifact, reported field, period, role, and available substantive validation identical. This design identifies a response to *reported portfolio volume*; it does not establish what happens when an actual person produces additional work under real production constraints.

Fully crossed factorial cells can become prohibitively numerous, particularly when discipline, evaluator expertise, and institutional status are also varied. A blocked fractional factorial or conjoint design could randomize several profile attributes while preserving sufficient overlap for estimation of prespecified effects. Hainmueller, Hopkins, and Yamamoto develop the potential-outcomes basis for randomized conjoint profiles and clarify relevant identification assumptions [10]. Any application here would have to specify whether the estimand is a response to a single profile attribute averaged over a particular distribution of other attributes or a conditional contrast within a defined institutional scenario. Interactions

between productivity and disclosure, or between disclosure and independent verification, would be especially important. These effects should be prespecified before inspecting outcome differences, because a large number of unplanned comparisons can produce misleadingly striking results.

A basic estimand for an evaluation outcome Y may be written as in Equation 19. Let $\mathcal{D}$ denote a prespecified joint distribution over evaluators o and other profile attributes X , and let n_H and n_L denote high and low reported output counts. The outcome may be a competence score, a binary offer of further examination, or a resource-related decision, but its scale and the target distribution $\mathcal{D}$ must be fixed before estimation.

$$\tau_Y(n_H, n_L; \mathcal{D}) = \mathbb{E}_{(o, X) \sim \mathcal{D}}[Y_o(n_H, X) - Y_o(n_L, X)]. \quad (19)$$

Equation 19 defines an average causal contrast for the stipulated joint distribution $\mathcal{D}$ only under the proposed randomized presentation and assumptions concerning consistency, assignment, positivity, and interference. If the experiment intentionally varies profile-attribute distributions across evaluator populations, the target distribution and corresponding estimand must be reported separately. It does not hold an actual researcher’s entire life history constant; its strength is narrower control of the information supplied to a particular evaluator. A negative effect on perceived competence at fixed validated quality would support the possibility of an unfavorable productivity signal. A null effect would limit that mechanism in the tested evaluator population and information setting. A positive effect would supply a counterexample to a generalized recognition-inversion claim.

To distinguish the role of AI disclosure from the role of output volume, the experiment should avoid replacing an informative contribution statement with an uninformative label. A categorical indication that an artifact was “AI-assisted” mixes different production processes, from proofreading to automated proof search. Experimental descriptions should instead identify which operations were performed by human participants and computational systems, what was checked, and who accepted responsibility. An additional source-label-only condition can estimate reactions to the label itself. Such a manipulation measures responses to wording under a controlled vignette, rather than a response to an actually altered production process. A study describing an identical process in inconsistent ways would also risk confusing participants; comprehension checks should confirm that the label and process description are interpreted as intended. Existing experimental studies find that AI labels can depress evaluations in some communication and creative-writing contexts, whereas other studies report weaker or absent effects for different outcomes [1, 20, 7]. Their outcomes and populations differ from scientific competence evaluation; they motivate a test rather than establish its result.

A crucial identification restriction concerns the review gate itself. If both the assigned portfolio description and latent evaluator interest affect whether an artifact is examined, analyzing competence ratings only among those who chose to examine it conditions on a post-treatment selection variable. Such conditioning can induce associations absent in the originally randomized evaluation population. One design should therefore elicit competence judgments from all randomized participants after a standardized minimum evidence presentation, separately from an optional-review task. Another may model examination as a measured intermediate outcome, with sensitivity analysis for unobserved judgments among unexamined cases. The resulting estimands concern different populations and should be reported separately; a favorable rating among selected reviewers cannot automatically stand for the effect of the intervention on the full evaluator population.

11.2 Disaggregated Evaluation Measures

A second requirement is measurement of the successive steps through which an artifact may or may not receive recognition. The distinction matters because an evaluator can decline to examine a portfolio while retaining a favorable view of the producer’s competence. Likewise, a reviewer can judge an artifact to be technically sound while doubting the producer’s independent understanding, or can find a creator capable without having budgetary authority to offer support. Treating these responses as one trust score would obscure the main mechanisms of the proposed theory.

Table 10: Candidate outcomes for disaggregating evaluative congestion, attribution, recognition, and resource-return mechanisms.

Outcome	Possible observation	Interpretive caution
Access and examination	Opening, review time, selected items	Attention is partly constrained by assigned workload
Artifact assessment	Blind validity, reproducibility, usefulness ratings	Technical standards vary by field and task
Competence attribution	Ability-to-explain or independently-extend judgment	May reflect documentation rather than stable ability
Contribution credibility	Confidence in role-specific responsibility claims	AI label alone rarely describes division of work
Cooperation or endorsement	Follow-up invitation, shortlist, recommendation	Selection and institutional incentives affect choices
Resource commitment	Bounded allocation, paid-task offer, support eligibility	Budgets and demand can dominate epistemic approval

A primary outcome battery should separately capture: the decision to open or review an artifact; actual time spent examining provided evidence; an assessed level of artifact correctness or usefulness; inferred task-relevant competence; the credibility of specified human contribution claims; willingness to initiate further discussion or verification; institutional endorsement; and hypothetical allocation of bounded resources. The use of time and click measures requires care: viewing duration can reflect confusion or interruption as well as attention, and a click does not indicate comprehension. Researchers should combine behavioral measures with narrowly specified judgment items and, where feasible, short tests of the evaluator’s understanding of the provided material.

The distinction between a focal artifact and an aggregate portfolio should remain explicit. Under a fixed review budget, the probability that any *particular* artifact is considered may decline as other submissions are added, while the number of favorable decisions over the whole portfolio may remain constant or increase. Section 8 established this difference algebraically. An experiment that reports only the fraction of items accepted could mistakenly infer diminished recognition from a stable number of positive judgments divided by a growing denominator. Conversely, a stable fraction could conceal a declining probability that work by a less connected producer enters the examination stage. Measurements must report denominators, selection rules, and the level of aggregation.

Trust also warrants multidimensional measurement. A respondent’s judgment that a particular result is reliable differs from their confidence that an author can independently explain a method, disclose limitations responsibly, or sustain a collaboration. The present paper should measure these components only insofar as they help discriminate the production–recognition mechanism; an integrated psychological model of human trust updating belongs to the companion inquiry on trajectory cognition. Similarly, willingness to pay for an artifact, propose a collaboration, or grant access to equipment should be recorded as distinct choices. These decisions depend on resources and obligations beyond belief in competence, as shown in Section 9.

The evidentiary treatment could be examined through an additional step in which a subset of evaluators receives independent validation after making an initial assessment. Researchers could compare revisions following genuinely diagnostic evidence with revisions following a prestige marker that adds no technical test result. A more favorable update after substantive validation may reflect appropriate learning; a response to a prestige marker alone may instead indicate the operation of a social heuristic. Yet this distinction cannot be resolved by an abstract status label without specifying what real-world reliability information the marker might reasonably convey. Any experimental status treatment should describe its informational content precisely.

Table 10 gives an integrated measurement scheme. Its purpose is to prevent causal explanations inferred from one outcome from being transferred without evidence to other stages of recognition.

The analytic plan should identify a small number of primary endpoints and treat the remaining mea-

2604 sures as secondary or exploratory. Latent trust factors can be investigated through psychometric models
2605 if measurement invariance is demonstrated across relevant evaluator groups; otherwise results should
2606 be reported at the item or carefully justified composite level. Missing judgments require attention as
2607 substantive outcomes: respondents who refuse to examine a portfolio or state that they cannot infer
2608 competence should not automatically be coded as distrustful. Their responses may reveal uncertainty, a
2609 capacity limit, or an appropriate refusal to infer beyond the evidence.

2610 11.3 Evaluator and Institutional Heterogeneity

2611 A third design would vary the institutional context of assessment. Scholarly journals, technical hiring
2612 committees, research funders, open-source maintainers, and public-interest organizations face different
2613 tasks and obligations. A journal primarily evaluates a submitted claim; an employer may need evidence
2614 of future competence and cooperation; a funder allocates limited resources subject to eligibility require-
2615 ments; a software maintainer may prioritize sustained reliability and community contribution. The same
2616 output history can therefore support different decisions without implying inconsistency or unfair treat-
2617 ment.

2618 A multilevel study could recruit evaluators from several roles and disciplines, stratifying assignment
2619 according to evaluation experience and familiarity with AI-mediated research. The same core artifact and
2620 profile treatment could be presented within each field after adapting technical criteria through pretested
2621 disciplinary rubrics. The design should not assume that a single specialized artifact is intelligible to
2622 every evaluator. Expert and nonexpert responses can be studied separately, and cross-field comparisons
2623 should specify which attributes and measurement scales are legitimately comparable. Oversampling less
2624 represented evaluator roles may be justified for mechanism discovery, with population reweighting only
2625 when the relevant sampling frame is known.

2626 The role of workload can be estimated by randomizing the number of portfolios assigned to evalua-
2627 tors, the time available, or the cost of obtaining verification evidence. Such interventions must distinguish
2628 induced congestion from changes in the substantive quality of submitted work. For example, with iden-
2629 tical underlying portfolios, one group might assess a manageable queue while another receives a larger
2630 queue under a fixed time budget. Recording not only the final decision but also the selection process
2631 allows estimation of the congestion route. Spillovers are possible: exposure to many AI-labeled profiles
2632 may change how later profiles are interpreted, even if their content is identical. The assignment unit, or-
2633 der effects, and exposure sequence should therefore be documented and, where appropriate, randomized
2634 or modeled as part of the treatment.

2635 Institutional status and prior acquaintance deserve special caution. A prestigious affiliation may carry
2636 informative signals about earlier screening and available resources; it can also become a source of unwar-
2637 ranted deference. A structured professional exchange may create valid evidence of understanding, while
2638 informal contact may give some applicants advantages unrelated to task performance. Tests of these
2639 possibilities can compare role-relevant technical interactions with minimal status-only cues, or examine
2640 assessments made before and after a structured question-and-answer session. These comparisons could
2641 help determine whether an intervention improves *evidence quality* or primarily shifts attention and so-
2642 cial standing. The full theory of heterogeneous relation events, human observers, and power-conditioned
2643 trust belongs to Papers II–IV; the present experiment uses these distinctions only where they affect the
2644 institutional recognition outcome.

2645 A particularly informative test compares two institutions with different evaluation rules but compara-
2646 ble decision tasks. One may allow detailed contribution evidence and independent technical discussion,
2647 while another relies chiefly on dossier screening and conventional output indicators. Observational com-
2648 parison alone cannot attribute different outcomes to the rules because applicant pools, resources, and
2649 institutional objectives may differ. A coordinated experiment with shared standardized fictional profiles,
2650 or a staggered policy trial if ethically and administratively feasible, could supply more discriminating
2651 evidence. Even then, implementation fidelity and evaluator learning would require direct measurement.
2652 Evidence that a procedure works for a particular kind of task would justify a domain-specific design

recommendation, rather than a claim that one model of recognition should govern every field.

11.4 Observational Research and Identification

Controlled profiles isolate particular informational mechanisms, but they cannot by themselves show whether the same processes materially affect actual research careers. A complementary research program would therefore examine longitudinal evaluation settings where artifacts, outcomes, and changes in assessment procedures can be observed lawfully and reliably. Potential settings include opt-in research-software contribution studies, public calls for proposals with archived evaluation criteria, openly documented project maintenance processes, and anonymized institutional review data under appropriate agreements. The main objective would be to track pathways from production and submission through examination, recognition, collaboration, and material commitment, while making the boundaries of accessible information explicit.

Selection at each stage is a central identification threat. A visible sample of published papers excludes work never submitted, papers rejected without full review, and contributors who lacked resources to complete an output. Likewise, contributions accepted into a public software repository may already have passed social and technical screening. If analysis is restricted to recognized outputs, it may systematically omit producers most affected by limited visibility. Researchers should specify the population at risk, recover denominators whenever possible, and report the proportion of decisions for which records are unavailable. Missing access and refusal decisions may be outcomes of interest rather than merely data-cleaning problems.

Repeated observations could relate a producer's output rate to later recognition while controlling for validated artifact quality, field, relevant opportunity, access to AI, available resources, and prior institutional standing. Yet many of these variables change in response to earlier recognition; conditioning on post-treatment attributes can therefore distort the estimated effect. A credible observational study needs a temporally specified causal question and an explicit description of potential confounding and selection. Within-producer comparisons can reduce stable differences in underlying competence but cannot remove time-varying shocks, changes in applicant strategy, or institution-specific exposure. Administrative data without evidence of independent evaluation cannot establish whether distrust, congestion, or resource scarcity explains an unfavorable outcome.

Changes in review policy might provide useful quasi-experimental variation, especially where the policy has a documented introduction date and affects a known subset of cases. Difference-in-differences, interrupted time-series, or event-study strategies could be considered if assumptions about comparison trends, anticipation, concurrent reforms, and composition changes are defensible. A policy introduced in response to a sudden increase in submissions is particularly vulnerable to endogeneity. The existence of a policy change should not itself be treated as a natural experiment. Researchers should examine pre-intervention trends, inspect simultaneous changes in applicant pools and funding, and present sensitivity analyses for unobserved disturbances. Methodological principles for randomized field designs and interpretive caution are discussed by Gerber and Green [9]; experimental transfer to actual institutional settings would likewise require consent, feasibility, and protection against unfair treatment.

The strongest evidence for recognition–return decoupling would combine credible evaluation records with subsequent resource outcomes. For example, a publicly validated scientific tool may receive endorsements without generating a paid maintenance role, while another tool may secure funding through a separate institutional support program. Such cases illustrate possible mechanisms, but neither one establishes a causal effect of recognition on income. Study designs would need to document budgets, eligibility, demand for the work, competing project priorities, and opportunity costs. Outcome measures should distinguish actual paid contracts or continuing institutional support from stated willingness to cooperate. Section 9 supplies the conceptual categories, while a more general theory of relational value circulation remains reserved for the fifth paper.

Table 11: Mechanism-specific observations and results that would challenge proposed explanations.

Proposed mechanism	Discriminating comparison	Result limiting that explanation
Output-count discount	Same validated focal artifact, varied reported volume	Stable or increased competence attribution at adequate precision
AI-label discount	Same process disclosure and evidence, varied source label	No material label response within prespecified interval
Review congestion	Same candidate set, experimentally varied review burden	No change in examination opportunities or delay
Provenance response	Same artifact, additional relevant inspectable records	No improvement in calibration or reduction in uncertainty
Recognition–return gap	Favorable validated judgments and later funding outcomes	Resource commitments consistently track recognition under known constraints

11.5 Alternative Explanations and Potential Falsifiers

A useful empirical program should be capable of undermining the paper’s preferred explanation. The recognition paradox is a conditional mechanism, not a diagnosis that can be applied whenever a prolific researcher receives a poor outcome. At least five competing possibilities must be assessed. First, a decline in average artifact quality may accompany expanded output in some circumstances. Second, substantial AI involvement may genuinely reduce the evidentiary value of an artifact for assessing an independently performed skill. Third, a larger portfolio may encounter lower per-item examination rates because review capacity is limited, even when competence judgments remain unchanged. Fourth, producers may create work for which there is little current institutional demand or funding. Fifth, status, discipline, language, or preexisting networks may govern visibility independently of productivity or AI disclosure.

These alternatives generate different observations. A quality-based explanation predicts that independently validated quality accounts for an adverse evaluation response. An attribution-based explanation predicts that quality can remain fixed while role-specific contribution uncertainty alters inferred competence, potentially with stronger effects when provenance is absent. A congestion-based explanation predicts changes in examination rates following workload variation, with little necessary effect on competence judgments conditional on examination. A resource-constraint explanation predicts favorable evaluations alongside absent or insufficient material commitments. A pure-label explanation predicts adverse reactions to an AI label even when the described process and available validation evidence are unchanged. Each is compatible with different forms of institutional response.

Table 11 summarizes discriminating tests. The term *falsifier* is used here in a bounded sense: a well-powered null result for a prespecified contrast in a defined setting limits that proposed mechanism in that setting, while still leaving open its operation elsewhere. One cannot empirically disprove an unrestricted claim that recognition inversion is possible in some conceivable circumstances; the scientifically meaningful claims are those with a specified population, task, treatment, and outcome.

Several interpretive cautions follow. An evaluator may properly reduce confidence when the evidence no longer supports a claim about unaided performance; this is different from concluding that AI-assisted work has low quality. A null average effect can conceal opposing subgroup effects, although subgroup claims require appropriate multiplicity controls and replication. Apparent benefits of added provenance may arise because evaluators infer effort from the documentation itself, rather than because the records resolve the intended attribution question. Experiments should therefore include checks of what respondents believe the manipulation conveyed, along with preregistered analyses and independent replication where feasible. When evidence is genuinely insufficient, a respondent’s refusal to form a definite competence judgment may indicate epistemic restraint rather than exclusion.

A further challenge concerns external validity. Creative-writing evaluation, social-media sharing, scientific peer review, technical hiring, and software-maintenance collaboration invoke different standards

of quality and responsibility. Findings in one domain should motivate comparisons in another without being treated as equivalent outcomes. The aim is to identify which institutional and informational conditions produce recognition inversion, rather than to assign one uniform psychological disposition to all evaluators. Evidence of careful acceptance, legitimate rejection, and well-calibrated uncertainty is as important as evidence of adverse discounts.

11.6 Ethical and Methodological Constraints

Research on trust and evaluation can itself change the opportunity structure of the people studied. An experiment in a real hiring or funding process should not manipulate consequential decisions or withhold material information in ways that unfairly disadvantage applicants. Early tests can use synthetic profiles and simulated allocation tasks in which no actual person's livelihood depends on experimental assignment. Where naturalistic institutional settings are studied, independent ethics review, informed participation as required, data minimization, confidential treatment of contribution histories, and procedures for institutional accountability should guide the design. Deception, when considered, requires a specific ethical rationale, proportionality to research value, and applicable approval; it should not be adopted merely to increase realism.

Power and sample planning should be outcome- and design-specific. Factorial interactions generally require substantially more information than simple main-effect comparisons, and repeated evaluations from the same respondent are statistically dependent. Pre-study simulations could examine plausible effect sizes, evaluator clustering, measurement noise, and the effect of uneven cell sizes on interval precision. A prespecified equivalence margin is needed before interpreting a null difference as evidence that two conditions are practically similar. Researchers should describe allocation, exclusions, missing-data handling, stopping criteria, and planned analysis in an accessible preregistration or equivalent protocol. These procedures improve interpretability but cannot overcome invalid measures or sampling frames.

Transparent materials should be balanced against intellectual property, confidential research, and individual privacy. Synthetic portfolios may be released with explicit labeling, codebooks, and randomization scripts. Real production histories may contain collaborators' contributions, proprietary datasets, correspondence, or information about institutional decisions, and thus may require limited-access audit rather than unrestricted publication. Provenance can strengthen legibility while also exposing contributors to new burdens and risks. This ethical tension is substantive to the research question: a procedure intended to improve recognition should be evaluated for the work and vulnerabilities it imposes on creators.

Finally, the proposed studies should avoid reproducing a presumed hierarchy between "human-only" and "AI-assisted" work. The relevant questions concern observable artifact quality, the distribution of responsibility, and the fit between evidence and the decision being made. Human-only production is not automatically reliable, and computational assistance does not independently guarantee correctness. Results should be reported in terms of the specific process disclosures and evaluations tested. Claims of institutional unfairness require evidence about criteria, consistency, alternatives, and context; favorable or unfavorable individual evaluations are not by themselves sufficient.

The proposed research program therefore proceeds through distinct observation gates. Controlled comparisons can identify responses to output volume, AI disclosure, and verified contribution evidence; review-all and selection arms can distinguish attention from assessment; institutional studies can examine later resource decisions. An unfavorable outcome warrants explanation rather than automatic classification as a recognition paradox. Section 12 considers the implications and limits of that inquiry.

12 Discussion and Scope of Generalization

The paper has separated the supply of intellectual artifacts, their selection for examination, inferences about contribution and competence, institutional endorsement, and resource-bearing commitments. These stages may adjust on different timescales when production technologies change. This discussion

draws out the implications for institutional design, epistemic justice, and the sustainability of public knowledge, while specifying the boundaries of generalization and the interfaces with companion inquiries. The formal model illustrates possible adverse responses under defined assumptions; the empirical designs in Section 11 are necessary to establish their occurrence and extent.

12.1 Institutional Adaptation across Timescales

The temporal structure of productive change provides one possible explanation for tensions between new forms of intellectual work and established assessment practices. A generative system can be introduced into an existing research workflow over a comparatively short interval. The same institution may require longer processes to revise hiring criteria, redistribute expert review time, authorize new contribution records, or create a budget for maintaining widely used research software. These different rates of adjustment should be analyzed in relation to particular organizational settings rather than treated as attributes of technology and society in the abstract. The transformation of production is itself uneven: additional computational capacity can accelerate some tasks while leaving fieldwork, experimental validation, interpretive judgment, and responsible deployment dependent on slower or less scalable processes [25, 11].

An institutional mismatch can emerge when an evaluation rule remains stable while the mechanism producing its observed input has changed. Consider a convention that treats the number of completed outputs within a specified period as a rough indicator of individual competence. Its usefulness depends partly on prior regularities connecting time, accessible resources, quality, and participation. When AI assistance changes the distribution of production costs, the meaning of the same reported count can become less determinate. Such a change does not by itself render the earlier convention irrational: some settings may retain useful correlations between sustained output and research competence. The evaluative question is whether the convention remains calibrated to the particular outputs, tasks, and populations for which it is used. A review rule can be outdated for one research practice and informative for another at the same institutional moment.

The time required for adjustment is also a function of the institutional object. Screening procedures can sometimes be revised by editorial policy. Norms of independent verification may develop through professional societies or collaborative projects. Promotion rules and employment classifications may require collective bargaining, formal governance, or changes to funding arrangements. Funding cycles can continue to constrain resource allocation even after a research contribution has been positively evaluated. The language of a uniform institutional “lag” is therefore too coarse unless it identifies which practice, which productive change, and which comparison interval are relevant. Institutional continuity may also preserve valuable safeguards. Deliberate review, cautious attribution, and requirements for reproducibility can protect knowledge quality in periods of rapid output growth [17, 22].

Table 12 identifies analytically distinct sites of adjustment and proposes observable indicators. The rows represent institutional functions with potentially different rates and constraints, not a developmental sequence every organization is expected to follow.

Comparison across these functions requires attention to the time structure of the underlying trajectories. Researchers can accumulate validated contributions before an institution creates a category in which those contributions can be assessed. Conversely, a new award category can appear before enough expert practice has developed to evaluate applicants reliably. A resource program can accept a project at one point yet terminate support before its maintenance obligations have ended. In each case, the order and duration of institutional events can affect the possibility of recognition and the conditions of subsequent work. A longitudinal design should therefore record decisions and their effective dates, together with changes in productive methods and the availability of external validation, rather than assume that calendar time alone is an explanatory variable.

The possibility of adjustment also makes recognition inversion historically revisable. If a negative response to high reported output arises because reviewers lack credible evidence about contribution or cannot afford substantive examination, changes in evidence interfaces or reviewer capacity may weaken

Table 12: Institutional functions relevant to adjustment under changing knowledge-production conditions. The indicators are proposals for comparative research rather than measured rates of adaptation.

Institutional function	Adjustment object	Candidate observation
Submission and discovery	Intake, indexing, screening, and access to review	Queue size; triage reasons; examination rates
Evidence and verification	Standards for validity, provenance, and human contribution	Audit requirements; reproduction rates; verification burden
Credentialing and career evaluation	Appointment, promotion, and task-relevant competence criteria	Rubrics; accepted evidence; appeal opportunities
Recognition and cooperation	Endorsement, collaborative assessment, and responsibility allocation	Follow-up decisions; joint verification; role descriptions
Resource return and maintenance	Employment, grants, contracts, and infrastructure support	Accessible funding; compensation; maintenance continuity

that effect. If high output instead correlates with genuinely inadequate verification, a more discriminating assessment procedure could appropriately sustain some unfavorable judgments. Neither possibility entails a predetermined convergence toward greater recognition. Institutional adaptation is a series of contested choices involving epistemic standards, resources, and authority. Its effects require examination within the relevant field and the particular producer population.

12.2 Epistemic Justice and Evaluative Responsibility

The evaluative distinction at the center of the paper also has a normative dimension. An investigator whose work is produced with AI assistance may face uncertainty about the degree to which a final artifact demonstrates particular human capabilities. Competent evaluation can recognize that uncertainty and request evidence relevant to the decision being made. A different concern arises when an evaluator assigns an adverse competence judgment solely from an unspecified AI label, an unfamiliar rate of output, or an institutional affiliation, without examining available and relevant evidence. The two responses should remain distinguishable. The former can be an appropriate exercise of epistemic responsibility; the latter may create an avoidable credibility disadvantage under some conditions. A claim of injustice requires further contextual and normative argument beyond the observation of a negative judgment.

The appropriate evidentiary standard depends on the purpose of evaluation. A journal that assesses the correctness of a technical claim may reasonably prioritize reproducibility and argument validity. An employer considering a position requiring independent problem solving may request a live technical explanation or a supervised task. A funder may also consider feasibility, public relevance, and the applicant's ability to maintain a project under specified obligations. None of these procedures can be reduced to one universal measure of diligence. Responsible scrutiny specifies the relevant capability, acceptable sources of evidence, conditions of access, and opportunity for an answer or correction. It also distinguishes uncertainty about contribution from demonstrated defects in the work. The proposals for diversified and qualitative research assessment advanced by DORA and CoARA offer relevant institutional starting points, although their adoption would not by itself resolve the AI-specific identification problems examined here [6, 4].

A further concern is that the burden of becoming assessable is unequally distributed. The ability to produce detailed experiment logs, independently validated benchmarks, long-term version records, or comprehensive contribution documentation depends on time, technical facilities, and administrative support. A well-resourced laboratory may incorporate these activities into its normal division of labor; an individual contributor may have to perform them alongside production and subsistence work. A rule requiring exhaustive proof of contribution from all applicants may appear formally uniform while having sharply different practical costs. On the other hand, eliminating evidence requirements entirely would expose collaborators and the public to avoidable uncertainty. An evaluative system has reason to seek *proportionate evidentiary legibility*: documentation and verification requirements calibrated to the

stakes, available alternatives, and actual inferential question.

Purpose-relative provenance is one possible means of reducing unnecessary burden. A research software contribution might be evaluated through a documented design decision, an inspectable test, and a role-specific explanation of the work, without requiring surveillance of every private thought or interaction with a model. A clinical or safety-critical tool will generally call for more extensive process and validation evidence. Contribution taxonomies and interoperable provenance descriptions provide useful vocabulary for such judgments [18, 15]. Their adequacy depends on implementation: a record is evidence of a described process only insofar as its origin, completeness, and relation to the asserted contribution can be responsibly evaluated. Recorded activity is not equivalent to technical correctness, and a missing record is not conclusive proof that a contribution was absent.

Epistemic responsibility extends to evaluators and organizations. Institutions choose how much reviewer labor to allocate, which kinds of accomplishments their forms admit, which appeals are possible, and whether unconventional evidence receives substantive consideration. Procedures are more contestable when applicants can learn which claim was evaluated, what evidence was decisive, and how factual errors can be corrected. The aim is the possibility of reasoned revision, including revision that confirms a previous rejection. Transparency has practical limits when confidential information, proprietary methods, or the privacy of third parties is involved. Procedures may therefore need trusted intermediaries, secure access to verification results, or bounded explanations that disclose evaluative grounds while protecting legitimate interests.

Recognition systems should also be examined for the unequal effects of status. Familiar institutions can supply genuinely useful information about prior selection, supervision, or reliability. Prestige can also serve as a shortcut that displaces the evaluation of an unfamiliar contributor's demonstrated work. The relevant normative question is whether the weight assigned to institutional signals is proportionate to their evidentiary value for the particular decision. Establishing that relation requires comparison with task-relevant evidence rather than an assumption that prestige is uniformly invalid or universally informative. The empirical designs proposed in Section 11 could isolate some such effects, although a complete account of the power relations governing recognition belongs to a separate inquiry.

12.3 Sustainability of Public Knowledge Production

The recognition paradox is especially consequential where productive activity creates public or widely shared knowledge infrastructure. A mathematical method, an open research-software library, a reproducible workflow, or a carefully documented dataset may enable later work by many users who never enter a financial transaction with its original contributors. Public availability can therefore expand downstream productive possibilities while leaving the producer's own material return uncertain. This observation does not imply that every publicly accessible artifact has high epistemic value, nor that each contributor thereby acquires an unconditional claim on collective funds. It identifies an institutional coordination question: how should an evaluative and resource-allocation system identify useful contributions and sustain the work required to keep them usable?

Open scientific infrastructure typically involves work beyond initial release. Maintenance may include correcting errors, updating dependencies, documenting compatibility, preserving data integrity, answering technically substantive questions, and revising tools when the underlying field changes. These obligations are distributed across different periods and people. Conventional output measures can recognize a new publication or release while giving limited visibility to the continuing activity that preserves its usefulness. Research on the sustainability of research software and the recommendations of public open-science institutions identify maintenance, infrastructure, and incentive arrangements as continuing challenges [2, 14, 24]. The present paper extends that concern by asking whether AI-mediated expansion of initial output may enlarge the stock of artifacts requiring evaluation and maintenance without generating a corresponding expansion in support.

This possibility generates multiple, partly competing dynamics. Greater productive capacity may make it cheaper to create high-quality tools, automate some maintenance activities, and enable new par-

2916 ticipants to contribute. It may also increase the number of overlapping packages, incompatible interfaces,
2917 unsupported prototypes, and claims that require expert checking. The effect on overall public knowledge
2918 capacity depends on the reliability, interoperability, documentation, adoption, and maintenance of the re-
2919 sulting outputs. A simple count of open repositories or preprints cannot settle whether the infrastructure
2920 available to a community has improved. A more informative longitudinal study would assess indepen-
2921 dent usability, repair rates, continuity of maintainers, integration into actual workflows, and whether
2922 users can reproduce substantive outputs.

2923 The allocation of returns presents an additional distinction between *public benefit* and *private liveli-*
2924 *hood*. Widespread downstream use can justify institutional interest in supporting a resource without
2925 establishing a direct channel from each use to compensation. Conversely, a paid researcher may be able
2926 to sustain valuable public contributions because employment income is provided independently of the
2927 specific audience for each release. Different arrangements, including salaried institutional stewardship,
2928 competitively awarded project grants, compensated maintenance, commissioned verification, and shared
2929 support through consortia, distribute responsibility and decision authority in different ways. Section 9
2930 treated these as resource-return interfaces. The present concern is their capacity to sustain an evolving
2931 productive process rather than to reward a single visible artifact.

2932 Where a creator's resources are limited, the cost of making work recognizable can compete directly
2933 with the cost of producing or maintaining it. Preparing tailored portfolios, securing independent reviews,
2934 meeting complex funding eligibility requirements, and attending professional events can each improve
2935 access in some settings while displacing necessary technical labor or ordinary subsistence. The tragedy
2936 discussed in Section 7 consequently has an institutional rather than merely psychological dimension. A
2937 productive system that relies indefinitely on uncompensated verification and maintenance can remain
2938 technically impressive yet expose its contributors to severe continuity risks. The risk is not confined to
2939 individual producers: knowledge users may also lose access when a valuable but unsupported project
2940 ceases to be maintained.

2941 An institutional response must accommodate both stewardship and contestability. Long-term support
2942 may improve continuity but can concentrate governance authority in funders or established institutions.
2943 Highly competitive short-term funding may distribute opportunities more widely in principle while creat-
2944 ing recurrent application costs and unstable maintenance. Distributed commons-based maintenance can
2945 broaden participation yet encounter coordination difficulties and uneven labor burdens. The appropriate
2946 arrangement depends on the public function of the resource, the dependence of its users, its risk profile,
2947 and available governance capacity. An assessment framework should therefore record whose work is
2948 recognized, who has authority to direct or discontinue it, and which resources are actually accessible
2949 to those expected to maintain the contribution. A complete theory of how such commitments circulate
2950 through relational structures remains a distinct research task.

2951 The status of the paper's contribution should be judged relative to established work on signaling, sci-
2952 entific evaluation, attribution, and research-assessment reform. Scarce attention, imperfect competence
2953 signals, and unequal institutional access are established problems. AI-mediated production motivates
2954 a particular joint configuration of these problems and permits a sharper, testable distinction between
2955 per-artifact examination, competence attribution, and resource return. The manuscript's claimed contri-
2956 bution is therefore an integrated conditional framework and a set of discriminating comparisons; it does
2957 not claim priority for the underlying existence of signaling, review congestion, or institutional inequal-
2958 ity. Establishing a historically distinctive AI-mediated mechanism would require comparisons across
2959 technologies, disciplines, and institutional adaptation periods.

2960 12.4 Interfaces with the Companion Papers

2961 The scope of the recognition paradox becomes clearer when situated within a series of neighboring ques-
2962 tions. The present paper takes AI-mediated production as its principal setting and investigates the insti-
2963 tutional interpretation of outputs, contributions, competence, and resource return. It treats changes in
2964 productive arrangements as relevant explanatory conditions and develops testable hypotheses concern-

ing evaluation and recognition. Its mechanisms do not provide a general theory of evolving relational trajectories, human cognition, power, or value circulation. Those objects warrant independent theoretical treatment because their causal structures and appropriate methods may differ from the institutional evaluation problem addressed here.

A second inquiry would examine heterogeneous events within evolving relational trajectories. An encounter with an unfamiliar kind of actor, institution, practice, or evidentiary interface can alter the accessibility of earlier work and the set of observers capable of examining it. This process can occur outside AI-mediated production and may have positive, neutral, or adverse consequences for trust. The structural study should distinguish the occurrence of events from their visibility to particular observers and should avoid equating novelty, informational entropy, and evidence supporting confidence. The present paper invokes relational encounters only as candidate institutional conditions under which substantive assessment might become possible. It leaves a full formal account of event heterogeneity and trajectory recognition to that separate study.

A third inquiry concerns human recognition of trajectories. Even when the same evidence is accessible, observers can differ in their attention, memory, prior beliefs, inferential heuristics, and interpretive categories. The production–recognition analysis necessarily simplifies these processes by representing observable judgments and conditional response functions. A cognitive theory would specify how observers form representations of distributed events, update expectations, and distinguish competence, integrity, and reliability. It could also investigate cognitive biases, selective attention, and retrospective interpretation of unchanged records. Such processes may mediate measured evaluation outcomes without being identified by the institutional model itself. The empirical designs in Section 11 offer possible observations, while the cognitive explanation remains an independent task.

A fourth inquiry would investigate the power structure of relational trajectories and its influence on epistemic access and trust. Review authority, control over information, eligibility rules, reputational hierarchy, and the capacity to authorize or withhold cooperation may condition whose contributions can be observed and validated. The present paper discusses unequal entry costs and evaluative authority because they affect recognition and sustainability, yet it does not derive a general model of power–trust coevolution. In particular, greater institutional visibility should not be treated as proof of epistemic merit, and power relations need not operate exclusively through distortion. Responsible institutional authority may provide independent verification and accountability as well as create exclusions.

A fifth inquiry would study the circulation of trajectories through relational currency. Recommendations, professional endorsements, cooperative commitments, recognized contributions, and other relational credentials may move between heterogeneous institutions under different conversion conditions. Their circulation cannot be identified with the movement of money, nor assumed to have a common numerical unit. Section 9 demonstrated why recognition and actual resources are distinct analytical objects. A general theory of their circulation, conversion, dissipation, and governance exceeds the parsimonious scope of the present paper. Maintaining these boundaries permits connections between the studies without assuming equivalence between their formalisms or prematurely resolving unresolved questions in any one domain.

12.5 Open Theoretical and Institutional Questions

Several questions remain unresolved within the present analysis. The first concerns the prevalence and shape of recognition inversion. The formal mechanism is possible under particular assumptions, but existing studies of disclosure effects or review congestion do not establish a general relation between the productivity of a scientific researcher and perceived competence. The size and direction of any such relation may differ by output type, evaluator experience, the declared role of AI, the availability of independent checking, and the institution’s decision rule. Controlled studies should identify the specific outcome whose response is being estimated. A decline in the fraction of artifacts receiving review is different from a decline in the evaluator’s posterior judgment of the producer’s ability or a reduction in total material support.

The second question concerns the boundary between a justified information discount and an unsupported credibility penalty. When the same artifact can arise through several distinct human–AI contribution histories, uncertainty about a particular person’s contribution may be appropriate. Evidence about role-specific decisions, independent explanation, and responsibility can narrow that uncertainty, although no documentation regime can guarantee complete access to an individual’s competence. Conversely, qualitative judgments can remain biased even when an elaborate portfolio is available. Empirical work should compare the diagnostic performance of alternative evidence arrangements, their documentation costs, and their susceptibility to strategic presentation. Evaluation reform has an opportunity cost of its own: extensive verification demands can reduce participation, while weak verification can transfer risks to users and the public.

A third question concerns the relative importance of evaluative capacity and its distribution. Expanding aggregate review time may improve overall examination opportunities yet leave systematic exclusion intact if the added capacity is directed toward already visible or institutionally favored portfolios. More efficient machine-assisted screening could free human expertise for substantive analysis, or it could amplify selection errors when deployed through opaque rules. Tests of review reform should measure whose work is sampled, what types of errors are detected, how disagreements are handled, and how much labor each procedure requires. Its success cannot be inferred from processing speed alone. The burden of rigorous assessment must also be budgeted; otherwise a reform aimed at improving fairness may depend on further uncompensated work.

A fourth question concerns the persistence of recognition–return decoupling. A positive judgment about the quality of a contribution does not determine whether any institution needs, can afford, or is authorized to support its creator. The mechanisms of employment, research funding, procurement, and collective maintenance differ in eligibility, duration, and obligations. Recognition may be a relevant input to these mechanisms without being a sufficient cause of material return. Understanding sustainable knowledge production will therefore require data on subsequent cooperation, accessible income, infrastructure use, and maintenance commitments, together with evidence concerning public benefits that ordinary financial transactions may not capture.

Finally, the historical and comparative range of the framework remains open. Disciplines differ in what counts as a contribution, how verification is conducted, whether output can be evaluated independently of its process, and how professional competence is demonstrated. Language, institutional access, public-resource constraints, and changing professional expectations may alter the costs of recognition. Some forms of AI-mediated production may make valid work easier to recognize by producing clearer documentation or more reproducible tests. Others may introduce ambiguity or accelerate the spread of errors. A credible general account will need to accommodate these divergent trajectories rather than extrapolate from a single kind of artifact, evaluator, or institutional setting.

The resulting institutional question concerns how to maintain informative assessment as productive arrangements change. Responsiveness to new evidence, proportionate documentation, accountable review, and support for continuing work may improve different stages of that process. Their relevance varies with disciplinary standards, resource constraints, and evaluative purposes. The conclusion states the paper’s contribution within these limits.

13 Conclusion

AI-mediated production can increase the supply of intellectual artifacts while changing the evidence available for evaluating their creators. This paper has investigated a conditional paradox: higher observable productivity may coincide with reduced access to examination, less favorable judgments of task-specific competence, or weaker institutional recognition even when relevant capabilities remain stable. These outcomes have different mechanisms and should be measured separately. The analysis combines conceptual distinctions, an illustrative formal model, institutional alternatives, and empirical research designs; it makes no claim to have estimated the prevalence of recognition inversion. The expression *curse of diligence* names both the proposed adverse-recognition possibility and the choices that

may arise when productive capacity expands faster than evaluative and resource-support arrangements.

The first contribution is an analytical separation of objects that are often combined in judgments of productive achievement. Output quantity records a count or rate of artifacts under a specified definition; epistemic quality concerns the support, validity, and relevant uses of those artifacts; competence concerns a contributor's capacity to undertake particular tasks; and contribution concerns activities and judgments exercised in a particular production process. Evaluative attention, trust, institutional endorsement, adoption, and material return are additional outcomes that may depend on different evidence and decision rules. Sections 2 and 5 established why a finished artifact can support a quality judgment without uniquely revealing the activities that produced it. The same ambiguity can arise for valuable work and defective work alike. Credible evaluation therefore requires attention to the claim being assessed and to the kind of evidence capable of supporting that claim.

The second contribution concerns the organization of evaluation under expanded production. As Section 4 argued, greater submission and discovery volumes may increase demands on scarce review time, specialist expertise, and the institutions that coordinate examination. The term *output inflation* was used in a restricted sense to describe potential pressure on the informational and attentional value of output signals. It does not assert a decline in the intrinsic worth of each contribution, nor does it establish that an increase in artifacts produces a uniform change in quality. An evaluative bottleneck may reduce the likelihood that a particular artifact receives serious examination while leaving an evaluator's judgment of its producer unchanged. This distinction matters for interpretation and intervention: increasing review capacity, changing selection rules, and improving the organization of evidence address different difficulties from those associated with competence attribution.

A third contribution is the conditional recognition-inversion hypothesis developed in Section 6. Production volume can retain a favorable signaling function when evaluators have reasons to associate sustained output with reliable contributions. Under other conditions, altered production costs, opaque divisions of labor, or poorly calibrated inference rules may make unusually high output a source of suspicion. The parsimonious model in Section 8 shows how a declining competence-attribution score is possible under explicit assumptions; it also distinguishes that result from declining per-artifact recognition caused by congestion. The illustrative derivations permit increasing, stable, and decreasing outcomes as assumptions change. Accordingly, a finding that a highly productive contributor receives fewer favorable responses cannot on its own establish that evaluators have unjustifiably discounted competence. Evidence about verified quality, task-specific contribution, evaluation exposure, and the relevant institutional criteria remains necessary.

The recognition paradox acquires a further dimension when creators face choices that carry costs in several directions. Section 7 examined how accelerated production may expand scientific and public opportunities while increasing the work required for validation, explanation, and maintenance. Selective restraint may improve a portfolio's examinability yet delay inquiries or public contributions. Extensive documentation may make a contribution more legible while diverting scarce time from inquiry, especially where research support is limited. Direct professional engagement can strengthen the grounds for substantive evaluation while requiring access, travel, coordination, and continued participation. These conflicts can persist even when the people involved exercise reasonable judgment. The tragedy therefore concerns the relation between feasible productive activity and the conditions under which that activity can be recognized and sustained; it should not be understood as a judgment on the virtue of diligence or on the intentions of particular evaluators.

This concern leads to a fourth distinction: recognition and material sustainability depend on partly separate institutional mechanisms. A contribution can be understood, independently verified, and regarded favorably while the person maintaining it has no corresponding employment opportunity, project funding, service contract, or maintenance support. Section 9 examined how budgets, eligibility rules, demand, ownership arrangements, and the organization of public knowledge can interrupt the passage from recognition to resources. Such interruptions bear particularly heavily on contributions whose public usefulness is compatible with weak direct revenue. The existence of epistemic value alone establishes neither a market price nor a claim to a specific allocation. Nevertheless, the resources required to con-

tinue research and maintain knowledge infrastructure remain a distinct object of institutional inquiry. An adequate account of productive sustainability must examine the costs of both making knowledge and keeping it usable.

The institutional responses considered in Section 10 accordingly pursue several related functions. Evidence about an artifact's claims should be proportionate to the risks and purposes of evaluation. Evidence about a contributor's understanding should address relevant tasks rather than rest exclusively on output counts, labels of AI participation, or inherited institutional prestige. Inspectable and appropriately bounded production trajectories may assist attribution, particularly when coupled with reproducible checks, opportunities for technical explanation, and independent review. The costs of providing such evidence also require scrutiny. Procedures that demand exhaustive records from contributors with unequal resources may reproduce the exclusion they seek to remedy. Contestable decisions, accessible routes to substantive examination, and transparent distinctions between failure of proof and proof of failure can improve the responsibility of evaluators and the practical possibility of justified correction. Their effectiveness depends on implementation and should be assessed rather than presumed.

The empirical program in Section 11 is essential to this argument's scientific status. Controlled comparisons can examine how portfolio volume and AI-use information affect attention, artifact assessment, inferred competence, and willingness to cooperate when other features are held fixed. Workload interventions can isolate congestion from attribution; independent validation can distinguish quality differences from source-label effects; and longitudinal institutional studies can investigate the conditions connecting recognition with actual support. Evidence contrary to the proposed mechanisms has equal analytical importance. If evaluators appropriately revise their judgments in light of demonstrable defects, if increased output improves recognition after quality is controlled, or if negative attribution effects disappear under realistic conditions, the scope of recognition inversion must be narrowed accordingly. The model's parameters and assumed relations have no established population estimates in this manuscript. Nor has the paper reported original experimental, field, or causal findings.

The broader theoretical question concerns institutional adjustment under changing production conditions. Evaluative rules embody historically developed expectations about time, capability, authorship, scarcity, and the organization of expert labor. Some of those expectations may become less diagnostic when productive mechanisms change; others remain necessary safeguards for public knowledge. Institutions adapt at different rates, and the effects of adjustment depend on domain-specific risks, reviewer capacity, governance, and the distribution of supporting resources. Reform therefore requires discriminating among procedures that responsibly preserve rigor, procedures that inadequately represent new forms of contribution, and procedures whose demands have become disproportionately costly. The recognition paradox identifies a possible source of institutional misalignment; it provides neither a universal diagnosis of distrust nor a comprehensive institutional blueprint.

Several complementary questions remain beyond this paper's scope. The informational and relational consequences of heterogeneous events in evolving trajectories, the cognitive processes through which observers represent and remember those trajectories, the power relations governing access to evidence and authority, and the circulation of relational recognition across institutional settings each require more detailed inquiry. They are connected to the present argument because recognition occurs through historically formed relations, yet the mechanisms developed here can be examined without presupposing a complete theory of trajectory cognition, power, or relational currency. Maintaining these boundaries enables the production-recognition hypothesis to be tested independently and permits the accompanying research to develop its own methods and counterexamples.

The strongest remaining theoretical challenge is to determine when a negative volume response is a warranted probabilistic inference, an avoidable selection effect, or a demonstrably miscalibrated discount. The framework itself does not decide this question. Its use requires task-specific competence definitions, reference populations, and evaluation standards, together with designs that observe the review-selection gate and retain results inconsistent with the proposed paradox. The argument is most informative when it identifies a separable mechanism that can be challenged by counterevidence.

Greater productive capacity can enlarge the range of possible intellectual contributions while leav-

3165 ing their examination, attribution, and sustainability contingent on social and institutional arrangements.
3166 Where an adverse response to output occurs, responsible analysis requires identifying whether its source
3167 is warranted inference, selective access to review, or poorly calibrated evaluation. Where valuable work
3168 receives recognition but lacks sustainable resources, a separate institutional explanation is needed. The
3169 paper offers a framework for making these questions distinguishable and empirically answerable. Their
3170 relative importance across disciplines and production arrangements remains to be established.

DRAFT

References

- [1] Sacha Altay and Fabrizio Gilardi. “People Are Skeptical of Headlines Labeled as AI-Generated, Even If True or Human-Made, Because They Assume Full AI Automation”. In: *PNAS Nexus* 3.10, pgae403 (2024). DOI: 10.1093/pnasnexus/pgae403.
- [2] Hartwig Anzt et al. “An Environment for Sustainable Research Software in Germany and Beyond: Current State, Open Challenges, and Call for Action”. In: *F1000Research* 9, 295 (2020). Version 2. DOI: 10.12688/f1000research.23224.2.
- [3] Erik Brynjolfsson, Danielle Li, and Lindsey Raymond. “Generative AI at Work”. In: *The Quarterly Journal of Economics* 140.2 (2025), pp. 889–942. DOI: 10.1093/qje/qjae044.
- [4] Coalition for Advancing Research Assessment. *The Agreement on Reforming Research Assessment*. Original agreement was released in July 2022; official CoARA web edition. July 20, 2022. URL: <https://www.coara.org/agreement/the-agreement-full-text/> (visited on 10/25/2026).
- [5] COPE Council. *Authorship and AI Tools: COPE Position*. Position statement; the page lists an earlier review date as well as an August 2024 version date. Aug. 20, 2024. DOI: 10.24318/cCVRZBms. (Visited on 10/25/2026).
- [6] Declaration on Research Assessment. *Frequently Asked Questions*. Official DORA explanation of responsible research assessment, qualitative evaluation, and diverse outputs. URL: <https://sfdora.org/faq/> (visited on 10/25/2026).
- [7] Isabel O. Gallegos et al. “Labeling Messages as AI-Generated Does Not Reduce Their Persuasive Effects”. In: *PNAS Nexus* 5.2, pgag008 (2026). DOI: 10.1093/pnasnexus/pgag008.
- [8] Claudine Gartenberg et al. “More Versus Better: Artificial Intelligence, Incentives, and the Emerging Crisis in Peer Review”. In: *Organization Science* 37.3 (2026), pp. 795–812. DOI: 10.1287/orsc.2026.ed.v37.n3.
- [9] Alan S. Gerber and Donald P. Green. *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton & Company, 2012.
- [10] Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. “Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments”. In: *Political Analysis* 22.1 (2014), pp. 1–30. DOI: 10.1093/pan/mp024.
- [11] Björn Hanneke et al. “Academic Publishing in the Age of Generative AI”. In: *Business & Information Systems Engineering* 68 (2026), pp. 729–737. DOI: 10.1007/s12599-026-01013-0.
- [12] Qian Yue Hao et al. “Artificial Intelligence Tools Expand Scientists’ Impact but Contract Science’s Focus”. In: *Nature* 649 (2026), pp. 1237–1243. DOI: 10.1038/s41586-025-09922-y.
- [13] International Committee of Medical Journal Editors. *Recommendations: Defining the Role of Authors and Contributors*. Section II.A.4: Artificial Intelligence (AI)-Assisted Technology. 2026. URL: <https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html> (visited on 10/25/2026).
- [14] Eric A. Jensen and Daniel S. Katz. “Evolving Funding Strategies for Research Software: Insights from an International Survey of Research Funders”. In: *PLOS ONE* 20.8, e0329833 (Aug. 21, 2025). DOI: 10.1371/journal.pone.0329833.
- [15] PROV-O: *The PROV Ontology*. W3C Recommendation. W3C Recommendation; editors listed on the specification, with additional working-group contributors. World Wide Web Consortium, Apr. 30, 2013. URL: <https://www.w3.org/TR/prov-o/> (visited on 10/25/2026).
- [16] National Academies of Sciences, Engineering, and Medicine. *Open Science by Design: Realizing a Vision for 21st Century Research*. Washington, DC: The National Academies Press, 2018. DOI: 10.17226/25116.
- [17] National Academies of Sciences, Engineering, and Medicine. *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press, 2019. DOI: 10.17226/25303.
- [18] National Information Standards Organization. *ANSI/NISO Z39.104-2022: CRediT, Contributor Roles Taxonomy*. National Information Standards Organization. Feb. 8, 2022. DOI: 10.3789/ansi.niso.z39.104-2022.
- [19] Shakked Noy and Whitney Zhang. “Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence”. In: *Science* 381.6654 (2023), pp. 187–192. DOI: 10.1126/science.adh2586.

- 3222 [20] Manav Raj, Justin M. Berg, and Rob Seamans. “The Artificial Intelligence Disclosure Penalty: Humans
3223 Persistently Devalue AI-Generated Creative Writing”. In: *Journal of Experimental Psychology: General*
3224 155.4 (2026), pp. 896–915. DOI: 10.1037/xge0001889.
- 3225 [21] Michael Spence. “Job Market Signaling”. In: *The Quarterly Journal of Economics* 87.3 (1973), pp. 355–
3226 374. DOI: 10.2307/1882010.
- 3227 [22] Springer Nature. *Editor AI Use and Assessment Guidance*. Official editorial policy and risk-based guidance;
3228 undated online resource. URL: [https://www.springernature.com/jp/policies/editorial-](https://www.springernature.com/jp/policies/editorial-policies/ai-editor-guidance)
3229 [policies/ai-editor-guidance](https://www.springernature.com/jp/policies/editorial-policies/ai-editor-guidance) (visited on 10/25/2026).
- 3230 [23] Elham Tabassi. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. Na-
3231 tional Institute of Standards and Technology, 2023. DOI: 10.6028/NIST.AI.100-1.
- 3232 [24] UNESCO. *Recommendation on Open Science*. Nov. 23, 2021. URL: [https://www.unesco.org/en/](https://www.unesco.org/en/legal-affairs/recommendation-open-science)
3233 [legal-affairs/recommendation-open-science](https://www.unesco.org/en/legal-affairs/recommendation-open-science) (visited on 10/25/2026).
- 3234 [25] Hanchen Wang et al. “Scientific Discovery in the Age of Artificial Intelligence”. In: *Nature* 620 (2023),
3235 pp. 47–60. DOI: 10.1038/s41586-023-06221-2.

Research Ethics and Source Provenance

Research status. Sections 1–13 constitute a complete revised theoretical manuscript awaiting independent scholarly review. The manuscript reports no original empirical study or institution-specific causal findings. Its formal proposition follows algebraically from explicitly stipulated assumptions; the illustrative model has not been empirically calibrated or validated. All illustrative cases are hypothetical and non-identifying.

Reference policy. Sections 1–13 use checked external sources recorded in the accompanying bibliography and verification ledger. Any additional references require verification before insertion.

Scope. The paper investigates conditional mechanisms and competing explanations. Its recognition hypotheses remain open to empirical and formal revision.