Generative Verificationalism: A Paradigm Shift in the Humanities and Social Sciences Based on AIGC

Shiyao Zhang, Yang Shen, Runzhi Yuan
Tsinghua University, The University of Sydney
Haidian District, Beijing, China - Lanqiyang Road, Tsinghua Road
13523099777@163.com

Abstract

Although the humanities and social sciences are rich in profound interpretive insights, they face an ongoing crisis of verifiability, which limits their interoperability with computational sciences. Traditional methodologies that rely on informal concepts and qualitative evaluations struggle to produce knowledge that is falsifiable and reproducible. The rise of agentic AI demands a fundamental shift in our research paradigm, which we term and define here as "Executable Humanities": a novel methodological framework whose core principle is that the validity of a theory must be demonstrated through its computational constructability. Within this paradigm, the primary aim of inquiry shifts from interpretation to construction, with AI agents serving as executors of formalized theories to generate testable artifacts and simulations. We decompose this paradigm into three foundational postulates: the Formalization Postulate, which asserts that meaningful concepts can be translated into computational language; the Generative Verification Postulate, which treats construction as the ultimate form of validation; and the Predictive Dynamics Postulate, which maintains that mature theories must be capable of modeling the evolution of systems. We employ the RCL unification theory not as an established law, but as an illustrative example to show how these postulates can be architected into a coherent system. This paper does not seek empirical validation of a specific model; rather, it aims to establish and advocate a new methodological foundation, and to invite the global scientific community to jointly inaugurate a new era of humanities and social sciences that are more rigorous, transparent, and computationally grounded.

1 Introduction

The humanities and social sciences, with their profound interpretive insights into the human condition, constitute an indispensable half of our understanding of the world. Yet within the modern scientific system, they have long faced a fundamental challenge—what we call the Verifiability Crisis. Sayre (2018) discusses the pervasive "reproducibility crisis" in science and its manifestation within the humanities. This crisis has created a formidable gulf between the wisdom of the humanities and the computational foundations of the natural sciences, limiting both the pace and the scope of knowledge accumulation.

It manifests in three interrelated maladies. First, the non-formalization of concepts: humanities and social science research often relies on heuristic but vaguely defined "soft" concepts, such as cultural capital, social identity, or historical consciousness. These concepts lack precise mathematical formulations and operational measurement methods, causing discussions based on them to devolve into endless textual interpretation rather than objective verification. Second, the irreproducibility of processes: research practices depend heavily on scholars' personal experience and intellectual intuition, resulting in opaque analytical pathways that are difficult for other researchers to independently replicate. As a result, the evaluation of knowledge often rests on trust in scholarly authority rather than on methodological reliability. Third, the non-falsifiability of conclusions: traditional paradigms tend to privilege post hoc explanations of events, but rarely generate ex ante predictions that can be quantitatively tested. A theory that cannot be rigorously falsified is, in the philosophical sense of science, of questionable validity.

With the rise of Agentic AI, this crisis has become more urgent than ever. Whereas earlier AI technologies (such as data mining or topic modeling) served primarily as "research assistants" for data processing, agentic AI possesses the ability to autonomously understand, plan, and execute complex tasks. This raises a critical question: if we cannot translate our theories into instructions that AI can comprehend and execute, then the most valuable insights of the humanities and social sciences will remain locked inside the "black box" of natural language, unable to fully leverage the transformative potential of this intelligence revolution.

Therefore, this paper argues that what is needed is not merely a new tool, but a profound paradigm shift in methodology. We introduce and define here the notion of Executable Humanities: a novel methodological framework designed to systematically resolve the aforementioned crisis. Its core principle is simple yet radical: the validity of a theory must be demonstrated through its computational constructability.

The purpose of this paper is not to empirically validate any single model, but rather to provide a systematic theoretical articulation of this new paradigm. We decompose it into three foundational postulates—Formalization, Generative Verification, and Predictive Dynamics—and demonstrate how they can help build a more rigorous, transparent, and creative future for humanities and social science research in the age of AI.

2 Paradigm Definition: Executable Humanities

Having established that the "verifiability crisis" constitutes a fundamental challenge for the humanities and social sciences, we now formally introduce a new methodological paradigm designed to systematically address this challenge: Executable Humanities. This paradigm is not merely a technical enhancement of existing research methods, but rather a reshaping of the very philosophy of inquiry. It calls for a fundamental redefinition of "understanding" as a core scholarly concept. Generative Verificationalism is a methodological philosophy that asserts the scientific validity of a theory is ultimately demonstrated by whether it can be used to construct an AI agent. This agent generates predictive phenomena or artifacts that can be empirically falsified in the real world, thereby testing and refining the theory.

2.1 From Interpretive Epistemology to Constructivist Epistemology (Core Principle)

The epistemological foundation of traditional humanities and social sciences has largely been interpretive. It holds that understanding is achieved through hermeneutic analysis of texts, thick description of cases, and discursive reasoning. This path undoubtedly yields profound insights, but its validity often relies on the intellectual authority of the interpreter rather than on an objective and reproducible process of verification. To transcend this limitation, we argue that the concept of "understanding" must be given a more rigorous and constructivist standard.

The philosophical cornerstone of the Executable Humanities paradigm is a constructivist epistemology, whose core principle is: to construct is to understand. This principle is not meant to diminish the value of interpretation, but rather to place it within a more demanding framework of verification. The internal logic proceeds as follows: if one claims to deeply understand a phenomenon—whether it be the emergence of a particular artistic style or the dynamics of a social movement—then the mere ability to provide eloquent descriptions is insufficient. A higher-order understanding requires researchers to formalize the phenomenon into its core constituent elements, the rules that govern their interactions, and the initial conditions that trigger its evolution. This represents an intellectual leap from "narrative telling" to "systematic blueprinting." As Dovetail Team (2023) emphasizes, the operationalization of concepts is crucial for transforming narrative insights into measurable data.

The ultimate and most unforgiving test of such a "blueprint" lies in its constructability—that is, in the ability to build, from scratch, a computational system capable of simulating, reproducing, or even autonomously generating the phenomenon in question. This resonates with physicist Richard Feynman's famous dictum: "What I cannot create, I do not understand." When a theory is translated into executable code, all of its ambiguities and tacit assumptions are forced into the open, for computational processes do not tolerate vagueness. A theory that resists formalization and construction inevitably calls its internal logical soundness into question. For example, Chen et al. (2024) developed a computational method for formalizing qualitative coding, illustrating how a theory, once implemented in code, reveals hidden assumptions and logical ambiguities.

Thus, the Executable Humanities paradigm requires a transformation in the researcher's core identity: from an "interpreter" into a "constructor." The ultimate goal of scholarship is thereby reshaped—not merely to compose a refined commentary on the world, but to build miniature systems capable of generating "working models of the world."

2.2 The New Role of AI Agents: As Dynamic Carriers of Theory

Grounded in a constructivist epistemology, the role of artificial intelligence (AI) agents in the research process has undergone a fundamental transformation. In past research practices, AI mostly functioned as a passive research assistant: a powerful efficiency tool applied to tasks such as data mining, pattern recognition, or information retrieval. For example, Jackson et al. (2017) systematically introduced agent-based modeling methods in social psychology, illustrating how early AI was positioned primarily as an assistant rather than as a true carrier of theory. In such applications, AI faithfully executed human-predefined instructions, but it did not "internalize" the theoretical constructs under investigation.

In contrast, within the framework of Executable Humanities, agentic AI becomes a dynamic carrier of theory and an active partner in execution. It is no longer merely an object of operation, but rather a "stage" on which thought experiments can unfold.

We define this new role as the "theory executor." This definition implies that a rigorously formalized theory can be encoded as the "operating system" of an AI agent—comprising its objective functions, environmental constraints, and decision heuristics. Thus, an AI agent is transformed from a generic tool into an actor infused with a specific "theoretical soul." Its mission is no longer to mechanically mimic surface-level human behavior, but rather to enact the inner logic of a given theory and autonomously explore its potential implications and consequences within its action space.

For example, a theory of tragic narrative could be formalized and embedded into a "playwright agent." In their work, da Costa et al. (2021) demonstrated how algorithmic information theory can be applied to computational creativity, showing that abstract artistic and literary theories can be transformed into computable rules for creative production. The task of such an agent would not be to imitate Shakespeare's style, but to execute the formalized rules of "tragedy theory" to generate entirely new scripts that nevertheless conform structurally to the core logic of tragedy. Similarly, a theory of opinion polarization could be encoded into a multi-agent simulation system. The research goal would then shift to observing whether agents, following only the interaction rules defined by "polarization theory," could spontaneously give rise to macro-level community fragmentation within simulated social networks.

The significance of this role transformation is profound. It marks a shift whereby AI no longer merely amplifies our ability to process historical data, but instead radically expands our capacity to test and develop theory itself.

2.3 Reconstructing the Research Process: From Linear Publication to Iterative Evolution

The redefinition of the role of AI agents necessarily requires a reconstruction of the entire research process. Traditional research in the humanities and social sciences is typically linear, culminating in a static academic publication as its endpoint. In contrast, the Executable Humanities paradigm introduces a more dynamic "construct–generate–iterate" cycle of research, inspired by contemporary practices in science and engineering.

Stage One: Formalization and Construction. Every inquiry begins with the most intellectually demanding step: translating a qualitative yet insightful theory into a rigorous formal language, and then constructing an executable AI agent or computational model. This construction process itself serves as a deep test of the theory's coherence and completeness.

Stage Two: Generation and Verification. Once constructed, the AI agent becomes a powerful "fact generator," capable of producing vast quantities of artifacts or simulated data under controlled conditions, consistent with the theory's expectations. Verification occurs on two levels: (1) internal verification, which tests whether the generated outputs are logically coherent and consistent with the theory's constraints; and (2) external verification, which compares the generated outputs with real-world data, expert judgments, or independent experimental results.

Stage Three: Dissemination and Refinement. The vitality of scientific knowledge lies in its circulation, critique, and cumulative development within the academic community. Thus, research outcomes should not be confined to static documents. We advocate an "open science" stance—open-sourcing our models, code, and datasets. AI agents themselves can be released as interactive tools. Through teaching, workshops, and online platforms, we invite global peers to test, challenge, and ultimately improve our work. Feedback from the community becomes the most valuable input driving the next, higher-level cycle of "formalization and construction."

This iterative cycle of construct–generate–iterate aims to transform humanities and social science research from a static enterprise culminating in publication into a dynamic, open, and cumulative scientific practice centered on the continual evolution of models.

3 The Three Pillars of the Paradigm: Core Postulates of Executability

Having defined the core principles and operational process of the Executable Humanities paradigm, we must now establish its logical foundation. This paradigm is not a loose methodological proposal, but one built upon three interrelated and progressively layered postulates. Together, these three postulates constitute a complete pathway for transforming qualitative insights into computable models, and they provide both the theoretical legitimacy and the practical feasibility for the role of AI agents as "theory executors."

3.1 The Formalization Postulate: The "Translatability" of Insight

Any clear, effective, and logically coherent theory in the humanities and social sciences must be translatable into one or more formal languages intelligible to AI, such as mathematical equations, logical rules, or computational algorithms.

This postulate is the logical starting point and prerequisite of the Executable Humanities paradigm. It boldly asserts that between the profound insights of humans and the computational intelligence of machines, there exists not an insurmountable ontological gulf, but rather a methodological "translation" challenge that can be overcome. We contend that when a theory appears "vague," "complex," or "non-computable," the cause lies not in the nature of the phenomena under study—such as culture, emotion, or power—being inherently resistant to formalization, but rather in the fact that the theory's internal logical structure has not yet been sufficiently explicated and rigorously defined.

"Formalization," therefore, is not merely a technical transcription, but a deep purification and intellectual stress test of theory itself. Klüver et al. (2003), in discussing computational sociology, noted that any real-world social system can, in principle, be completely modeled and simulated by a formal system equivalent to a universal Turing machine. This implies that there is no fundamental barrier to formalization. Formalization forces researchers out of the fog of descriptive language and compels them to confront the core skeleton of their theories. Concretely, this process requires three intellectual tasks: (1) precise definition of concepts, such as decomposing expansive terms like "social capital" into measurable variables—e.g., number of network nodes, tie strength, information flow rate, and trust transmission functions; (2) specification of relational hypotheses, such as rendering vague causal claims like "A significantly influences B" into computable functional relations (e.g., linear, exponential, or sigmoid curves), conditional probabilities, or strict logical entailments; and (3) delineation of boundary conditions, clarifying the contexts and parameter ranges in which a theory holds, and where it may fail.

A theory that cannot undergo this "translation" will often reveal its conceptual confusion, circular reasoning, or internal contradictions in the process. Conversely, a theory that can be successfully formalized demonstrates clarity, coherence, and structural robustness. Guest and Martin (2021) emphasize that computational modeling is itself the formal expression of theory, compelling researchers to articulate the precise components of their frameworks. This supports our claim that formalization helps researchers move beyond descriptive vagueness and face the core structure of theory, while also serving as a strong test of theoretical consistency. Thus, formalizability itself becomes a meta-standard for judging theoretical quality—it marks the first step in transforming personal insights into public knowledge. Through formalization, AI agents evolve from "laborers" that merely process data into "apprentices" capable of understanding and executing human thought. Once a theory has been successfully "translated," it ceases to be just a subjective belief in the scholar's mind or a hermeneutic object in a text; it is elevated to an objective knowledge entity that can be loaded, executed, and tested by machines. This lays the indispensable logical and technical foundation for the subsequent postulates of Generative Verification and Predictive Dynamics.

3.2 The Generative Verification Postulate: The "Proof Power" of Construction

The strictest and most profound verification of a theory does not come from its ability to fit or explain existing data, but from its capacity to guide an AI agent to successfully construct high-quality artifacts or systemic phenomena that match the theory's expectations.

This postulate directly challenges the dominant paradigm of verification in traditional humanities and social science research, which revolves around conformity tests. Whether through carefully selected case studies or complex statistical regressions, the essential logic remains testing how well a theory aligns with stock data. While such methods are necessary to scientific practice, their limitations have become increasingly evident, as they cannot fully rule out statistical ghosts such as multicollinearity, endogeneity, and overfitting to particular datasets.

By contrast, Generative Verification advances a more active and constructive standard of validation, shifting the focus from "explaining the past" to "creating the future." Its core logic is that a model capable of generating the expected phenomenon must, to a significant extent, have captured the causal mechanisms underlying that phenomenon, going beyond mere variable correlations. Because every action of the AI agent strictly follows the micro-rules encoded by the theory, the emergent macro-level phenomena can be seen as the logical entailments of those rules—amounting to a complete enactment of the causal chain.

Furthermore, the generative process allows us to conduct "stress tests" of theory. By systematically altering the initial conditions and key parameters of the model, we can observe the theory's performance under extreme or unprecedented scenarios. This makes it possible to identify a theory's "phase transition points" and "failure boundaries," thereby deepening our understanding of its applicability and vulnerabilities—insights unattainable by merely analyzing limited historical data.

Generative verification raises the standard from passive descriptive capacity to active constructive capacity. It no longer asks, "How well does your theory explain this existing world?" but instead poses a more exacting question: "According to your theory, can you construct a logically and phenomenologically convincing new world?" This provides a tool as powerful as a particle collider in physics: by "colliding" the internal rules of theory in a computational world, we can observe what new phenomena emerge. As Epstein (1999) famously declared, "If you didn't grow it, you didn't explain it." Our formulation here directly corresponds to Epstein's notion that only by "growing" macro-level patterns through agent-based models can one claim to have explained them. In this paradigm, construction itself is proof.

3.3 The Predictive Dynamics Postulate: The "Extrapolative Power" of Theory

A mature, fully formalized theory must inherently encode the internal dynamics that describe how the system under study evolves over time. Thus, a theory should not only generate static artifacts but also be capable of producing quantitatively testable predictions about the system's future states. This postulate sets the ultimate scientific goal for the Executable Humanities paradigm and completes its scientific closure. If formalization is the starting point of theory and generation is its process of validation, then prediction is the final measure of theoretical maturity.

The value of a theory lies not only in describing a system's static structure (e.g., a society's class hierarchy or a text's narrative framework), but also in explaining how that structure forms, persists, and changes. This postulate asserts that any valuable theory must contain an executable time-evolution operator. This operator may take many forms: a set of differential equations describing variable change, a Markov chain defining state transition probabilities, or an algorithm specifying multi-agent interaction protocols. Regardless of form, it must be capable of driving the system forward in time.

More importantly, such extrapolation must yield quantitative, falsifiable predictions, not vague statements like "the future may become more complex." An executable theory must be able to generate predictions such as: "Given intervention X, the polarization index of the target group will, within the next 30 days, rise from 0.5 to 0.7 ± 0.05 with 95% probability." Such predictions, along with preregistered confidence intervals and failure criteria, constitute the most severe test of a theory. Predictive success greatly enhances a theory's credibility, while predictive failure—if honestly documented and analyzed—provides invaluable input for its revision.

This advances humanities and social science theories from "static description" to "dynamic maturity." It marks a theory's transition beyond classification and explanation, toward grasping the causal laws underlying phenomena. It requires theories not only to answer "what" and "why," but also to answer, in a testable way, "what comes next." Hempel and Oppenheim (1948) likewise argued that valid causal explanations should serve as the basis for prediction. Today, the rise of computational social science has begun to overturn the traditional bias against prediction, urging researchers to place greater emphasis on predictive capacity. When a theory attains the ability to quantitatively forecast the future, it transcends mere post hoc explanation and achieves a higher level of scientific maturity. This is not only the ultimate marker of scientific progress, but also provides, for the first time, a solid and anticipatory foundation for social interventions and policy-making. This final piece of "extrapolative power" completes the logical closure of Executable Humanities as a comprehensive scientific paradigm.

4 Architecture Example: The RCL Unification Theory as an Illustrative Case

Having systematically articulated the three core postulates of the Executable Humanities paradigm—Formalization, Generative Verification, and Predictive Dynamics—we must address a key question: can these abstract philosophical principles be concretized into a coherent and operational theoretical architecture? If so, what would it look like? To answer this and to demonstrate how the three postulates can be instantiated in practice, we introduce the RCL (Rarity–Control-of-Entropy–Connectivity) Unification Theory as an illustrative example.

Before proceeding, we must clarify the purpose of this section. We present the RCL Unification Theory not as a fully empirically validated, definitive theory of cultural and social dynamics. On the contrary, we position it as a methodological concept prototype. Its core value lies not in the correctness of its present conclusions, but in its function as a thought experiment that clearly demonstrates the feasibility and concrete pathways for integrating our three postulates—translatability, constructability, and extrapolatability—into a unified, computable framework. RCL serves as a proof-of-concept that the paradigm shift we advocate is both intellectually and technically workable.

The RCL theory seeks to address a longstanding fundamental challenge in the humanities and social sciences: how to develop a unified understanding—and optimization—of the full lifecycle of an idea or work, from its creative genesis, to the formation of its intrinsic quality, to its social diffusion and impact.

4.1 How RCL Embodies the "Formalization Postulate": Computational Translation of Core Concepts

The Formalization Postulate requires that "soft" concepts in the humanities and social sciences be translated into languages intelligible to AI. RCL performs this translation by introducing three core computable dimensions that systematically formalize the highly abstract notions of creativity, quality, and propagation power.

From "creativity" to "rarity" (R). Traditionally, "creativity" or "originality" is subjective and elusive. RCL formalizes it as rarity on the basis of statistical learning. Park et al. (2020) propose an information-theoretic framework in which the novelty of a work can be quantified as the negative log of its generation probability. In a large generative model (e.g., an LLM), every possible idea corresponds to some generation probability in a "latent space of thought." Commonplace ideas occupy high-probability regions, whereas truly novel and surprising insights lie in low-probability "no man's lands." Hence, an idea's rarity score R can be defined as its negative log-likelihood, $-\log P(\text{idea})$. This definition turns "creativity" from a mystical notion into an objective, computable quantity grounded in probability and information theory.

From "quality" to "entropy control" (L). The "quality" or "texture" of a work is likewise a fuzzy evaluative notion. Drawing on information theory, RCL formalizes it as entropy control. It assumes that any information-bearing artifact (text, image, music) exhibits an information entropy HHH. Too little entropy implies high predictability, redundancy, and tedium; too much entropy implies disorder and unintelligibility. High-quality works typically reside within an entropy "golden window" [H_min, H_max], striking a subtle balance between structural predictability and informational novelty. This formalization provides a nonlinear, tunable control parameter for assessing intrinsic quality.

From "propagation power" to "connectivity." The influence of a work ultimately lies in the breadth and depth of its dissemination. Rajeh et al. (2021) found that in social networks, the most influential nodes are not the typical "stars," but rather the bridge nodes that connect otherwise separate communities. When information can traverse these bridges, it becomes far more likely to achieve cross-community diffusion. Building on network science, RCL formalizes this dimension as connectivity. Instead of treating society as a homogeneous "mass," it models it as a complex network G composed of distinct communities. A work's connectivity score C depends on whether its diffusion path can effectively activate the bridge nodes in the network, thereby enabling cross-community spread. This score can be quantified through network metrics such as betweenness centrality of the nodes along the path.

Through these three translations, RCL transforms a complex humanities problem into a computable, analyzable theoretical architecture defined in a multidimensional space—a concrete instantiation of the Formalization Postulate.

4.2 How RCL Embodies the "Generative Verification Postulate": An Objective Function as an AI Creation Engine

The Generative Verification Postulate requires that a theory possess constructive power. RCL operationalizes this by defining a unified objective function J(γ)that turns the theory into an execution engine guiding AI agents in creative generation.

This objective evaluates the overall value of any potential "creation–diffusion pathγ" . A simplified form is:

$$J(\gamma) = w\_r * R(\gamma) - w\_l * |H(\gamma) - H\_optimal| - w\_c * C\_cost(\gamma) - w\_risk * RISK(\gamma)$$

where w are the weights, R rewards rarity, H penalizes deviation from the optimal entropy band, C_cost represents connectivity costs, and RISK penalizes factual or ethical risks.

Elgammal et al. (2017) proposed Creative Adversarial Networks (CAN), which train AI to deviate from existing styles (novelty) while not straying too far (coherence/quality)—a concrete example that AI can generate under multi-objective criteria rather than mere mimicry. Each time an AI system generates outputs that meet such criteria, it furnishes a constructive validation of the underlying theory. In this framework, the AI agent is not a mimic, but a path-finder in a complex value landscape defined by RCL—searching, planning, and optimizing to maximize J(γ), thereby producing outputs that are highly original (high R), exhibit refined quality (small entropy deviation), possess strong diffusion potential (low connectivity cost), and maintain low risk.

Thus, RCL not only describes what constitutes a “good work,” but also provides an executable blueprint enabling AI to autonomously construct such works. Every successful generation is a strong, constructive test of RCL’s theoretical assumptions—the essence of the Generative Verification Postulate.

4.3 How RCL Embodies the “Predictive Dynamics Postulate”: A Simulator of Social Diffusion

The Predictive Dynamics Postulate requires extrapolative power. Following Bertotti et al. (2016), who extend the classic Bass diffusion model into a stochastic differential equation (akin to a Langevin equation) to simulate innovation or information diffusion on networks—with a deterministic “social force” term and a stochastic noise term—RCL incorporates dynamic equations to move from a static evaluation scheme to a system capable of simulating and predicting social phenomena.

For example, along the connectivity dimension, RCL does more than statically assess the bridging value of nodes. It can integrate dynamic equations from theoretical physics and epidemiology to simulate the spread of information over a social network G. An initial model can be approximated by a Langevin equation:

$$dX_t = F(X_t)dt + \sigma dW_t$$

where X_t denotes the coverage state of nodes over time, $F(X_t)$ is the deterministic driving force determined by the “opinion potential field” and network structure, and σdW_t represents stochastic perturbations.

By solving or simulating such equations, the RCL framework becomes a “future simulator”: given initial conditions (e.g., the seeding node of a message), it can generate quantitative predictions of the future diffusion curve, peak impact, and critical outbreak nodes. These predictions can be compared directly with real-world data, offering the most stringent—and most valuable—opportunities for falsification, thereby perfectly embodying the Predictive Dynamics Postulate.

The value of RCL lies in demonstrating that constructing a unified, computable, and verifiable theoretical framework for the humanities and social sciences is entirely possible in principle. It provides a robust first prototype for the broader methodological paradigm we advocate—one that future researchers can critique, refine, and surpass.

References

[1] Bertotti, M. L., Brunner, J., & Modanese, G. (2016). Innovation diffusion equations on correlated scale-free networks. Physics Letters A, 380(33), 2475–2479. https://doi.org/10.1016/j.physleta.2016.06.003

[2] Chen, Z., et al. (2024). A computational method for measuring open codes. arXiv. https://arxiv.org/html/2411.12142v1

[3] Da Costa, G. A. A. B., Dowe, D. L., & Hernández-Olivan, C. (2021). Computational creativity and aesthetics with algorithmic information theory. Entropy, 23(12), 1654. https://doi.org/10.3390/e23121654

[4] Dovetail Editorial Team. (2023, February 5). What is operationalization? definition & how-to. https://dovetail.com/research/operationalization/.

[5] Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). CAN: Creative adversarial networks, generating “art” by learning about styles and deviating from style norms. In Proceedings of the 8th International Conference on Computational Creativity (pp. 96–103).

[6] Epstein, J. M. (1999). Agent-based computational models and generative social science. Complexity, 4(5), 41–60. https://doi.org/10.1002/(SICI)1099-0526(1999)4:5<41::AID-CPLX9>3.0.CO;2-F

[7] Guest, O., & Martin, A. E. (2021). How computational modeling can force theory building in psychological science. Perspectives on Psychological Science, 16(4), 789–802. https://doi.org/10.1177/1745691620970585

[8] Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. Philosophy of Science, 15(2), 135–175. https://doi.org/10.1086/286983

[9] Jackson, J. C., Rand, D., Lewis, K., Norton, M. I., & Gray, K. (2017). Agent-based modeling: A guide for social psychologists. Social Psychological & Personality Science, 8(4), 387–395.

[10] Klüver, J., Stoica, C., & Schmidt, J. (2003). Formal models, social theory and computer simulations: Some methodical reflections. Journal of Artificial Societies and Social Simulation, 6(2), 8.

[11] Noailabs. (2025, March 13). Ai agents as humans // social experiments simulations. Medium. https://noailabs.medium.com/ai-agents-as-humans-social-experiments-simulations-5140553533cd

[12] Park, D., Nam, J., & Park, J. (2020). Novelty and influence of creative works, and quantifying patterns of advances based on probabilistic references networks. EPJ Data Science, 9(2), 2. https://doi.org/10.1140/epjds/s13688-019-0214-8

[13] Rajeh, S., Savonnet, M., Leclercq, E., et al. (2021). Characterizing the interactions between classical and community-aware centrality measures in complex networks. Scientific Reports, 11, 10088. https://doi.org/10.1038/s41598-021-89549-x

[14] Sayre, F., & Riegelman, A. 2018. The reproducibility crisis and academic libraries. College & Research Libraries. https://crl.acrl.org/index.php/crl/article/view/16846/18452.

Agents4Science AI Involvement Checklist

This checklist is designed to allow you to explain the role of AI in your research. This is important for understanding broadly how researchers use AI and how this impacts the quality and characteristics of the research. Do not remove the checklist! Papers not including the checklist will be desk rejected. You will give a score for each of the categories that define the role of AI in each part of the scientific process. The scores are as follows:

•[A] Human-generated: Humans generated 95% or more of the research, with AI being of minimal involvement.
•[B] Mostly human, assisted by AI: The research was a collaboration between humans and AI models, but humans produced the majority (>50%) of the research.
•[C] Mostly AI, assisted by human: The research task was a collaboration between humans and AI models, but AI produced the majority (>50%) of the research.
• [D] AI-generated: AI performed over 95% of the research. This may involve minimal 175
human involvement, such as prompting or high-level guidance during the research process, but the majority of the ideas and work came from the AI.

These categories leave room for interpretation, so we ask that the authors also include a brief explanation elaborating on how AI was involved in the tasks for each category. Please keep your explanation to less than 150 words.

IMPORTANT, please:
• Delete this instruction block, but keep the section heading “Agents4Science AI Involve- ment Checklist",
• Keep the checklist subsection headings, questions/answers and guidelines below.
• Do not modify the questions and only use the provided macros for your answers.

1.Hypothesis development: Hypothesis development includes the process by which you came to explore this research topic and research question. This can involve the background research performed by either researchers or by AI. This can also involve whether the idea was proposed by researchers or by AI.

Answer: [A]
Explanation: [The research topic ("Executable Humanities" and "The Crisis of Verifiability") and the core research questions were entirely proposed by the author. AI only provides limited assistance in literature organization and expression optimization, but the research hypotheses and problem framework were all generated manually.]

2.Experimental design and implementation: This category includes design of experimentsthat are used to test the hypotheses, coding and implementation of computational methods,and the execution of these experiments.

Answer: [A]

Explanation: [This paper does not involve empirical experiments or data-driven design. The methodological framework (such as the three core assumptions and the RCL unification theory) was independently constructed by the authors, and AI was not involved in the design or implementation process.]

3.Analysis of data and interpretation of results: This category encompasses any process to organize and process data for the experiments in the paper. It also includes interpretations of the results of the study.

Answer: [A]
Explanation: [This study does not involve data analysis or result interpretation. Its main contribution lies in the construction of theories and methodologies. Therefore, there is no role for AI in this aspect.]

4.Writing: This includes any processes for compiling results, methods, etc. into the final paper form. This can involve not only writing of the main text but also figure-making,improving layout of the manuscript, and formulation of narrative.

Answer: [B]
Explanation: [The main body of the paper (concept innovation, theoretical deduction, structural framework) was written by the author. AI provided assistance in some aspects such as language polishing, paragraph refinement, and logical connection, but the overall writing and narrative were led by the author.]

5.Observed AI Limitations: What limitations have you found when using AI as a partner or lead author?

Answer: [A]
Description: [During the process of using AI, its limitations were discovered: AI often tends to produce superficial summaries, lacks a deep understanding of complex philosophical arguments, and easily overlooks key methodological details. Without manual critical screening, it may lead to generalization of concepts or logical leaps. Therefore, in this article, AI is only used as a writing tool, not as a theoretical source.]

Agents4Science Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research,addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: Papers not including the checklist will be desk rejected. The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions.
For each question in the checklist:
• You should answer [Yes] , [No] , or [NA] .
• [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
• Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers and area chairs. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given. In general, answering "[No] " or "[NA] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:
• Delete this instruction block, but keep the section heading “Agents4Science Paper Checklist",
• Keep the checklist subsection headings, questions/answers and guidelines below.
• Do not modify the questions and only use the provided macros for your answers.

1. Claims
Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [Yes]
Justification: [The abstract and introduction sections accurately reflect the core contributions and scope of this paper. The paper proposes and defines a new paradigm called "executable humanities", elaborates on its three core axioms (formalization, generative verification, predictive dynamics), and provides illustrative examples through the RCL unification theory. The introduction clearly states that the objective of this article is to construct and advocate at the methodological level, rather than conducting empirical research. Therefore, it is consistent with the scope of contributions.]
Guidelines:
• The answer NA means that the abstract and introduction do not include the claims made in the paper.
• The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
• The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
• It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations
Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]
Justification: [This article acknowledges its own limitations: The RCL theory is defined as a "conceptual prototype" rather than an ultimate theory that has undergone complete empirical verification. At the same time, the article emphasizes that its contribution lies in the methodological and scientific philosophical aspects, rather than proposing specific algorithms that can be immediately applied or large-scale experiments. This means that its operability is limited in the short term, but it leaves room for future criticism, improvement, and application.]
Guidelines:
• The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
• The authors are encouraged to create a separate "Limitations" section in their paper.
• The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
• The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
• The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting.
• The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
• If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
• While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs
Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]
Justification: [This paper, at the theoretical level, presents three core postulates and elaborates on the corresponding assumptions one by one, such as: any meaningful concept can be formalized (formalization

postulate), verification should be based on generative construction (generative verification postulate), and the theory should possess the ability to predict future dynamics (predictive dynamics postulate). Although no mathematical theorems or strict proofs were provided, all the assumptions were clearly stated and explained in the text.]

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if 290 they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main ex-perimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [NA]

Justification: [This article does not include empirical experiments or data-driven research, so there is no issue of experimental reproducibility. The main contribution of the article lies in proposing a new methodological paradigm and philosophical framework, rather than experimental conclusions.]

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [NA]

Justification: [This study did not involve empirical experiments, datasets or specific codes, so there is no need for open data and code. Its contributions are mainly focused on theoretical and methodological construction.]

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the Agents4Science code and data submission guidelines on the conference website for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [NA]

Justification: [This article did not conduct experiments, therefore there are no training/testing details, parameter settings or optimizers, etc. of the experiments available for reporting.]

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.

• The full details can be provided either with the code, in appendix, or as supplemental material.

7.Experiment statistical significance
Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [NA]
Justification: [This article does not have an experimental section, so it does not involve statistical significance tests or reporting of error intervals.]
Guidelines:
• The answer NA means that the paper does not include experiments.
• The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
• The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, or overall run with given experimental conditions).

8. Experiments compute resources
Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [NA]
Justification: [This study does not rely on specific experimental or computational resources, therefore there is no need to report information such as hardware, running time or storage regarding the computing environment.]
Guidelines:
• The answer NA means that the paper does not include experiments.
• The paper should indicate the type of compute workers CPU or GPU, internal cluster,
or cloud provider, including relevant memory and storage.
• The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.

9. Code of ethics
Question: Does the research conducted in the paper conform, in every respect, with the Agents4Science Code of Ethics (see conference website)?

Answer: [Yes]
Justification: [This article strictly adheres to scientific research ethics. All discussions are based on publicly available academic materials and theoretical frameworks, and do not involve privacy data or studies of potentially sensitive populations, nor do they touch upon harmful uses (such as false information or surveillance).]
Guidelines:
• The answer NA means that the authors have not reviewed the Agents4Science Code of Ethics.
• If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.

10. Broader impacts
Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]
Justification: [In the conclusion and discussion sections of this article, the positive and potential negative impacts are emphasized. On one hand, it provides a more rigorous and transparent paradigm for humanities and social science research, which may enhance interdisciplinary collaboration and scientificity; on the other hand, it also highlights the risks of excessive formalization and excessive reliance on AI, which may undermine the interpretive tradition of the humanities. Therefore, the paper not only points out the potential benefits of society, but also acknowledges the possible negative effects.]
Guidelines:
• The answer NA means that there is no societal impact of the work performed.
• If the authors answer NA or No, they should explain why their work has no societal
impact or why the paper does not address societal impact.

• Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations,
privacy considerations, and security considerations.
• If there are negative societal impacts, the authors could also discuss possible mitigation 387strategies.