Beyond Prediction: A Neuro-Symbolic Architecture for Exploring High-Plausibility Futures

The Possibility Machine Paradigm: A Shift from Predictive Modeling to Combinatorial Exploration

The proposed 'Possibility Machine' represents a significant methodological departure from conventional Large Language Models (LLMs). Standard LLMs operate as predictive engines, trained on vast corpora of historical text to learn statistical patterns of language . Their core function is to predict the next most probable token in a sequence, effectively mimicking human language based on past exposure [62] [79] . This paradigm excels at tasks requiring stylistic coherence and summarization of known information but is fundamentally limited when tasked with generating genuinely novel insights, especially in domains like future oncology where the solution lies outside current datasets . The 'Possibility Machine' aims to transcend this limitation by shifting the focus from prediction to exploration. Instead of forecasting what is likely based on history, it seeks to explore the entire state space of possible linguistic combinations and then apply a rigorous filtering mechanism to identify configurations with high factual plausibility . This aligns directly with the research goal of generating speculative essays on futuristic breakthroughs, such as a cure for cancer, which necessitates venturing beyond existing knowledge .

This new paradigm is rooted in the concept of the Library of Babel, a theoretical library containing every possible combination of its characters . In this context, the Library represents the complete combinatorial search space of all potential written knowledge. The central challenge, therefore, is not generation itself, but the process of discerning valuable "needles" from the overwhelming "haystack" of meaningless gibberish, which constitutes over 99.999...% of the raw Library . The 'Possibility Machine' is conceived as the computational model capable of performing this discernment. It acknowledges the "combinatorial explosion" problem, which severely limits the length of text that can be practically processed—initial estimates suggest sentences up to eight words long, with optimizations potentially extending this to fifteen . This pragmatic constraint highlights that the project is not about brute-forcing through the entire infinite library, but about engineering a sophisticated sieve to find meaningful fragments within a manageable subspace.

A critical distinction must be made between the goals of factual plausibility and aesthetic coherence. While the user initially expressed interest in preserving "poetic nonsense" and "dream logic," the clarified objective prioritizes factual plausibility for generating scientifically grounded essays . This necessitates a filtering pipeline that moves beyond stylistic mimicry to incorporate semantic checks against established scientific principles and validated knowledge bases . The model is not merely seeking to write well; it is seeking to write correctly, with its output being constrained by the logical structure of reality as encoded in external knowledge systems. This transforms the model from a "fact machine" into a "possibility machine"—an explorer of the space of plausible truths rather than a generator of random text . The framing is inspired by the Mystic Seer from The Twilight Zone, a device that provides cryptic yet prophetic answers, suggesting the output should be presented as ranked possibilities of varying degrees of feasibility rather than definitive assertions .

The implementation plan addresses the scalability issue head-on by proposing synthetic generation instead of attempting to download the physically unmanageable Library of Babel . By using an API or a universal algorithm to generate specific coordinates, the system can create a massive but finite corpus of text fragments. These fragments would then pass through a multi-stage filtration process designed to progressively prune away noise and retain only the most promising candidates . This approach reframes the challenge from one of storage to one of computation, focusing resources on the intelligent filtering required to navigate the vast linguistic landscape. The ultimate aim is to create a system that understands the aesthetic of language without being tethered to reality, but whose creative outputs are ultimately guided and validated by a robust, knowledge-driven verification process . This positions the 'Possibility Machine' not as a replacement for scientific inquiry, but as a powerful tool to accelerate it by systematically exploring the space of potential discoveries.

The Multi-Stage Filtration Pipeline: From Linguistic Validity to Semantic Grounding

The success of the 'Possibility Machine' hinges entirely on the effectiveness of its multi-stage filtration pipeline, a systematic process designed to transform an overwhelmingly noisy combinatorial space into a curated set of semantically plausible text fragments .

This pipeline operates in three distinct tiers, each applying a different class of rules to progressively narrow down the search space from purely statistical regularities to meaningful scientific concepts. The first two stages establish a foundation of linguistic validity, ensuring that the remaining text is not only grammatically correct but also structurally plausible as a fragment of human language. The third stage introduces semantic grounding, leveraging external knowledge to validate the factual content of the filtered text.

The first stage of filtration is Shannon Entropy Thresholding. This method applies information theory to distinguish between different types of text based on their informational content. The formula for Shannon entropy, $H(X) = -\sum_{i=1}^{n} P(x_i) \log_b P(x_i)$, quantifies the unpredictability or randomness of a signal [8]. Purely random noise, such as static, exhibits maximum entropy because each character appears with equal probability. Conversely, highly repetitive sequences like "aaaa..." exhibit minimum entropy because the next character is completely predictable. Human language occupies a "sweet spot" of intermediate entropy, possessing enough structure to be meaningful but enough variation to convey novel information . By calculating the Shannon entropy of generated text blocks, the system can discard anything falling outside the empirically observed range for natural languages [4] [9]. This filter efficiently eliminates pure gibberish while preserving text that has the statistical signature of human communication, including "poetic nonsense" and "dream logic" that may still contain latent meaning . This technique is widely recognized in cognitive science and medical decision-making for formalizing uncertainty and has been applied in machine learning contexts to incentivize reasoning and improve model performance [4] [9] [11].

The second stage builds upon the first by introducing Markovian phonotactic filtering. While entropy analysis ensures a block of text has the right statistical properties, it does not guarantee that the constituent "words" are pronounceable or follow the sound patterns of any conceivable language. This stage addresses that gap. Using a simple N-gram model—a type of probabilistic language model trained on multiple world languages—the system can check if the sequences of characters conform to plausible phonotactic rules . For instance, a string containing no vowels for fifty consecutive characters would be discarded as linguistically implausible . This technique is supported by extensive research in computational linguistics, which focuses on modeling the morpho-phonological systems that build word forms from underlying units [36] [37]. Markov models have been successfully used to acquire morphological structure and combined with other methods like Latent Semantic Analysis for topic segmentation [38] [57]. A more advanced implementation could leverage Markov Logic Networks (MLNs), which fuse probabilistic graphical models with first-order logic, allowing for the representation of complex dependencies and exceptions in linguistic rules [47] [50]. This phonotactic filter acts as a crucial gatekeeper, ensuring that the vocabulary of the generated text is not just statistically valid but also structurally coherent, significantly reducing the set of invalid candidates before they reach the computationally intensive semantic analysis phase.

The third and most critical stage is Semantic Density Scoring, which bridges the gap between linguistic validity and semantic plausibility. At this point, the text has passed tests for statistical structure and phonotactic conformity, but it could still be nonsensical prose. To address this, the system employs lightweight embedding models, such as a distilled version of BERT, to analyze the semantic content of the text fragments . Each sentence or phrase is converted into a high-dimensional vector that captures its meaning. If the resulting vector is a "zero vector" or clusters far from any known concepts in the embedding space, it suggests the text is semantically inert and is discarded . This "semantic density" score measures whether the generated text maps to a recognizable part of human knowledge. This approach is validated by numerous studies demonstrating that LLMs implicitly model semantics to predict co-occurrence patterns [72] and that embeddings can effectively capture semantic similarity in neural activation spaces [46]. However, simply checking for a non-zero vector is insufficient. The true power of this stage emerges when these embeddings are used to query external knowledge systems, such as biomedical ontologies, to see if the concepts mentioned in the text are related in ways that are consistent with established scientific facts. This moves the evaluation from a binary pass/fail test to a nuanced assessment of plausibility, which can then be used to rank the final speculative essays.

Filter Stage	Primary Function	Key Method(s)	Rationale & Supporting Evidence
Linguistic Coherence	Discard text lacking statistical structure of language.	Shannon Entropy Calculation	Human language has a characteristic entropy level; this distinguishes it from pure noise (high entropy) and repetition (low entropy) [4] [8] [9].
Structural Plausibility	Ensure generated "words" follow plausible linguistic sound patterns.	N-gram / Markov Model Analysis	Checks for violations of phonotactic rules (e.g., excessive consonants), ensuring text is structurally plausible [38] [57].
Semantic Grounding	Verify that text fragments relate to meaningful concepts.	Lightweight Embedding Models (e.g., Distilled BERT)	Maps text to a vector space; discards "zero vectors" or text clustering far from known concepts, measuring semantic density [46] [72].

Integrating External Knowledge Systems: The Role of Ontologies and Neuro-Symbolic Reasoning

The transition from linguistically valid text fragments to scientifically plausible speculative essays requires a deep integration with external knowledge systems. The third stage of the filtration pipeline, Semantic Density Scoring, cannot succeed in isolation; it must be coupled with robust mechanisms for querying and reasoning over structured, domain-specific knowledge . The foundation of this integration rests on large-scale Biomedical Knowledge Graphs (KGs) and the use of neuro-symbolic architectures that combine the generative power of LLMs with the verifiability of symbolic logic. This fusion allows the 'Possibility Machine' to move beyond surface-level semantic matching and engage in genuine, high-assurance reasoning about the plausibility of its generated hypotheses [26] [49] .

Biomedical Knowledge Graphs, such as the Gene Ontology (GO), serve as the bedrock of this system by providing community-consensus views of biological entities and their relationships [60] . These KGs encode vast amounts of curated, structured knowledge, making them ideal for verifying the claims made within a generated essay. The iKraph project demonstrates the immense power of such systems; by constructing a massive KG from PubMed abstracts and augmenting it with data from public databases, researchers were able to perform probabilistic semantic reasoning to infer indirect causal links between entities [22] . In a real-time drug repurposing study for COVID-19, the iKraph system identified hundreds of candidate drugs per month, with a significant fraction of its early predictions later being validated as effective treatments [22] . This validates the premise that a well-constructed knowledge base can indeed guide discovery. However, the same study highlights a critical limitation of LLMs operating on unstructured text: they struggle with directional entailments and inconsistent knowledge retrieval [22] . This underscores the necessity of grounding the LLM's outputs in a structured KG to avoid hallucination and ensure factual accuracy.

To connect the free-form text generated by the 'Possibility Machine' to the structured predicates within a KG, specialized alignment and mapping pipelines are essential. Tools like GenOM and RELATE provide proven methodologies for this task. GenOM uses an LLM to generate textual definitions for ontology concepts, enriching their semantic representations to improve alignment across heterogeneous knowledge sources [24] . The RELATE pipeline offers a more direct path for the 'Possibility Machine'. It operates in three stages: preprocessing ontology predicates into searchable embeddings, retrieving candidate mappings based on semantic similarity, and using an LLM-based reranker to select the best match within the context of the source document [25]. This hybrid approach is particularly powerful because it leverages the broad semantic understanding of the embedding model for initial retrieval and the contextual awareness of the LLM for final validation. A key finding from the RELATE study is its strong capability in handling negation, correctly identifying underlying relationship types even when they are expressed with negative polarity [25]. This ability to process statements like "X does not affect Y" is not a niche feature but a fundamental requirement for generating scientifically accurate text, as it allows the model to rule out incorrect hypotheses and refine its conclusions.

Ultimately, the most advanced integration is achieved through neuro-symbolic architectures, which formally embed the LLM within a framework of verifiable symbolic reasoning. The LOGicalThought (LogT) architecture serves as a compelling blueprint for this approach [26]. LogT converts a natural language problem into a dual context of an ontologically-grounded symbolic graph and a machine-readable logic program. This grounding forces the LLM to adhere strictly to the rules and facts encoded in the ontology, producing more transparent and reliable reasoning traces. In evaluations across several benchmarks, LogT demonstrated significant improvements over standard prompting methods, particularly in handling complex logical constructs like negation, implication, and defeasible reasoning (+11.84% average improvement) [26]. Defeasible reasoning, which allows for rules to have exceptions, is crucial in domains like medicine and law, where general principles can be overridden by specific facts [26]. By adopting a neuro-symbolic framework, the 'Possibility Machine' can ensure that its speculative essays are not just grammatically correct and semantically dense, but also logically consistent with the foundational principles of the domain they describe. This architecture directly addresses the need for a "generator–verifier paradigm," where any hypothesis generator can be paired with a formal verifier to enhance the quality and reliability of scientific discovery [16].

Practical Implementation and Output Synthesis: From Synthetic Generation to Ranked Essays

The practical realization of the 'Possibility Machine' requires a concrete implementation plan that navigates the immense scale of the combinatorial search space while building a functional system capable of producing the desired output: ranked speculative essays on futuristic scientific breakthroughs . The strategy must be scalable, computationally feasible, and aligned with the core design principles of synthetic generation, progressive filtration, and knowledge-grounded synthesis. This involves creating a manageable corpus of text fragments, training the model to understand its unique identity as an "explorer," and developing a final synthesis engine that transforms filtered fragments into coherent, high-plausibility narratives.

The first step is to abandon the impractical notion of downloading or storing the entirety of the Library of Babel, which is too large to exist physically . Instead, the system will employ synthetic generation. A Python script can be engineered to generate a massive number of random text pages, perhaps a billion or more, by generating their corresponding coordinates algorithmically via an API . This approach treats the Library as a mathematical object that can be sampled from, rather than a physical archive to be possessed. These billions of generated pages form the raw input for the "Sieve" Layer, the first practical application of the multi-stage filtration pipeline. This layer runs the synthetic text through the initial filters: Shannon Entropy Thresholding to remove pure noise and Markovian Phonotactic Rules to discard linguistically incoherent strings . The result is a significantly smaller, yet still vast, corpus of text that possesses both statistical and structural plausibility. This pre-filtered corpus becomes the training and processing target for the subsequent, more computationally expensive stages of semantic analysis.

Once a manageable corpus is established, the Contrastive Learning component of the implementation plan comes into play. This technique is used to train the LLM to distinguish between "Pure Noise" and "Library Fragments" that have passed the initial linguistic filters . By presenting the model with pairs of examples—one from the noise category and one from the filtered fragment category—the model learns to internalize the subtle statistical and structural differences that define a "valuable" piece of text. This process is instrumental in imbuing the model with its core identity: that of an explorer navigating the Library of Babel, adept at recognizing the faint signatures of meaning amidst the noise . This training not only improves the model's ability to filter new text but also shapes its generative style, encouraging it to produce prose that carries the distinctive "vibe" of Borgesian, surrealist thought rather than generic, modern English . This framing is central to achieving the desired aesthetic of the Mystic Seer, which provides profound yet enigmatic answers .

The final stage is the synthesis of the filtered and ranked fragments into a coherent speculative essay. The 'Possibility Machine' will not produce a single essay but a ranked list of possibilities, each accompanied by a plausibility score . This score will be derived from the results of the semantic filtering stage. An essay's plausibility could be calculated based on several factors: the number of key concepts found in the biomedical knowledge graph, the strength and consistency of the relationships connecting those concepts, and the overall semantic density of the narrative. The synthesis engine would take the highest-ranked fragments and weave them together, using the LLM's generative capabilities to construct a fluent, persuasive essay that reads like a plausible scientific paper. The neuro-symbolic components, like the LogT architecture, would play a vital role here, ensuring that the connections made by the synthesizer are logically valid according to the underlying ontology [26]. The final output would thus be a carefully constructed narrative, representing a high-probability pathway through the combinatorial space of potential knowledge, offering humanity a glimpse into a possible future of scientific discovery.

Systemic Challenges and the Evolving Role of AI in Scientific Discovery

While the 'Possibility Machine' presents a compelling vision for accelerating scientific discovery, its development and deployment are fraught with systemic challenges that must be addressed. These hurdles span computational scalability, dependency on external knowledge, and the inherent risks of bias and error. Acknowledging these limitations is crucial for setting realistic expectations and guiding future research. Furthermore, positioning the 'Possibility Machine' within the broader context of AI in science reveals its potential not as a standalone oracle, but as a sophisticated collaborator in an evolving agentic research ecosystem.

The most immediate challenge is computational scalability. The initial synthetic generation of billions of text pages is resource-intensive, and the subsequent filtering stages present even greater demands. Running this massive corpus through a lightweight embedding model for semantic density scoring is a non-trivial task, and querying a large-scale knowledge graph like iKraph for every single fragment is computationally prohibitive at this scale [22]. The user's optimistic estimate of a 15-word limit on generated sentences may prove to be overly conservative once the full cost of semantic filtering is considered . Optimizing the efficiency of the entire pipeline—from faster embedding models to more efficient knowledge graph query algorithms—will be a primary area of research. Another major risk is the quality and completeness of the external knowledge bases upon which the system relies for validation [22]. Biomedical ontologies are constantly evolving, and knowledge extraction from literature can introduce errors or biases. The plausibility of the 'Possibility Machine's' output is therefore entirely contingent on the fidelity of its ground truth source. Any inaccuracies or gaps in the knowledge graph will be propagated into the generated essays, potentially leading to false leads.

Furthermore, the very nature of the model introduces subtle but significant risks. The process of defining "semantic density" and operationalizing the plausibility score remains a complex open problem. The system must also carefully navigate the tension between the desire for "poetic nonsense" and the strict goal of factual plausibility, ensuring that the generated essays maintain a scientific tone while being rigorously constrained by the knowledge base . There is also the risk of the model inadvertently learning and amplifying societal biases embedded within its training and knowledge sources [80]. Addressing these issues requires a multi-pronged approach, including rigorous validation against independent datasets, continuous monitoring of the knowledge base for updates and corrections, and the development of better metrics for quantifying plausibility, potentially drawing from information-theoretic measures like the Jensen-Shannon divergence [6].

In the broader landscape of AI-driven scientific discovery, the 'Possibility Machine' fits into a growing trend of moving from computational oracles to autonomous research partners [48]. Current research shows that LLMs are emerging as promising generators of scientific ideas, capable of producing coherent and sometimes factual hypotheses [13] [18]. However, their strength is often in augmentation rather than sole authority, especially in knowledge-intensive fields like biomedicine [22]. The 'Possibility Machine' formalizes this role by explicitly separating the generative function of the LLM from the verificatory function of the external knowledge system, adhering to a robust generator–verifier paradigm [16]. Its greatest contribution may lie not in providing definitive answers, but in systematically exploring the vast space of undiscovered scientific truths, highlighting the most plausible pathways for human researchers to investigate. By framing its output as a ranked list of speculative possibilities, it respects the complexity of scientific inquiry and avoids the pitfall of equating mimicry with genuine intelligence [83]. Ultimately, the 'Possibility Machine' represents a step towards a more collaborative future, where AI systems act as powerful sensemaking tools, helping scientists navigate the ever-expanding body of knowledge and accelerate the journey from question to discovery [40].

Reference

1. The 53rd Annual Meeting of the Association for https://aclanthology.org/P15-2.pdf
2. Computer Science https://www.arxiv.org/list/cs/new?skip=25&show=1000
3. (PDF) Deep Active Inference and Scene Construction https://www.researchgate.net/publication/350636922_Deep_Active_Inference_and_Scene_Construction
4. Generalized Information Theory Meets Human Cognition ... https://onlinelibrary.wiley.com/doi/10.1111/cogs.12613
5. Breaking through the classical Shannon entropy limit https://arxiv.org/pdf/2501.00612
6. A complex plausibility belief Jensen–Shannon divergence ... https://www.sciencedirect.com/science/article/abs/pii/S0952197624012144
7. A Semantic Generalization of Shannon's Information ... https://www.mdpi.com/1099-4300/27/5/461
8. This is IT: A Primer on Shannon's Entropy and Information https://telecom-paris.hal.science/hal-02287963/document
9. Generalized Information Theory Meets Human Cognition https://www.researchgate.net/publication/323570404_Generalized_Information_Theory_Meets_Human_Cognition_Introducing_a_Unified_Framework_to_Model_Uncertainty_and_Information_Search
10. ENTROPYLONG: EFFECTIVE LONG-CONTEXT TRAIN https://openreview.net/pdf/20b4c8bfa23b806c916a275bd86685ddf08e0480.pdf
11. Fully Unsupervised LLM Reasoning Incentivization https://arxiv.org/html/2504.05812v1
12. The Plausibility Trap: Using Probabilistic Engines for ... https://arxiv.org/html/2601.15130v1
13. Exploring the role of large language models in the scientific ... https://www.nature.com/articles/s44387-025-00019-5
14. SafeScientist: Enhancing AI Scientist Safety for Risk-Aware ... https://aclanthology.org/2025.emnlp-main.116.pdf
15. The role of LLMs in theory building https://www.sciencedirect.com/science/article/pii/S2590291125003456

16. The Need for Verification in AI-Driven Scientific Discovery https://arxiv.org/html/2509.01398v1
17. ICEAGE: Intelligent Contextual Exploration and Answer ... https://dl.acm.org/doi/10.1145/3733723.3733731
18. (PDF) Large Language Models for Scientific Idea Generation https://www.researchgate.net/publication/397521714_Large_Language_Models_for_Scientific_Idea_Generation_A_Creativity-Centered_Survey
19. Scientific Discovery - Stanford Encyclopedia of Philosophy https://plato.stanford.edu/archives/sum2019/entries/scientific-discovery/
20. The Evolution of AI: From Classical Machine Learning to ... http://ieeexplore.ieee.org/iel8/6287639/10820123/11202920.pdf
21. KEBLM: Knowledge-Enhanced Biomedical Language ... https://www.sciencedirect.com/science/article/pii/S1532046423001132
22. A comprehensive large scale biomedical knowledge graph ... https://pmc.ncbi.nlm.nih.gov/articles/PMC10760044/
23. The Short Text Matching Model Enhanced with Knowledge ... https://arxiv.org/abs/2304.03898
24. GenOM: Ontology Matching with Description Generation ... https://arxiv.org/abs/2508.10703
25. Relation Extraction in Biomedical Abstracts with LLMs and ... https://arxiv.org/html/2509.19057v2
26. Logic-Based Ontological Grounding of LLMs for High ... https://arxiv.org/html/2510.01530v1
27. Uncertainty-Informed Screening for Safer Solvents Used in the ... https://pubs.acs.org/doi/10.1021/acs.jcim.5c00612?ai=556&mi=0&af=R
28. Bayes-Entropy Collaborative Driven Agents for Research ... https://arxiv.org/pdf/2508.01746
29. Self-Evolution Learning for Discriminative Language Model ... https://aclanthology.org/2023.findings-acl.254.pdf
30. Large-Language-Model-Enabled Text Semantic ... https://www.mdpi.com/2076-3417/15/13/7227
31. Automated construction of inorganic materials databases ... https://journals.sagepub.com/doi/10.1177/14727978251352139
32. Active Inference AI Systems for Scientific Discovery https://arxiv.org/html/2506.21329v4

33. Sophimatics and 2D Complex Time to Mitigate ... https://www.mdpi.com/2076-3417/16/1/288
34. Proceedings of the 11th Meeting of ... https://aclanthology.org/W10-22.pdf
35. arXiv:2406.10885v1 [cs.CL] 16 Jun 2024 https://arxiv.org/pdf/2406.10885?
36. Quantitative and Computational Approaches to Phonology https://www.cambridge.org/core/elements/quantitative-and-computational-approaches-to-phonology/FEFA2295D874AF96D37972DC1BC1BBF5
37. Synthesizing theories of human language with Bayesian ... https://pmc.ncbi.nlm.nih.gov/articles/PMC9427767/
38. A Markov model for the acquisition of morphological structure https://www.researchgate.net/publication/229026678_A_Markov_model_for_the_acquisition_of_morphological_structure
39. Proceedings of the 12th Global Wordnet Conference https://aclanthology.org/2023.gwc-1.pdf
40. The Multifaceted Sensemaking Theory: A Systematic ... https://www.mdpi.com/2079-8954/11/3/145
41. Bacteria to AI: Human Futures with our Nonhuman Symbionts https://academic.oup.com/chicago-scholarship-online/book/61133/book-pdf/64294087/upso-9780226835983.pdf
42. (PDF) 'Speaking Brains' Collective Papers on AI, Neuro ... https://www.researchgate.net/publication/380461457_'Speaking_Brains'_Collective_Papers_on_AI_Neuro-circuity_and_Basal_Ganglia_Grammar
43. William Morgan - The Dragon, the Witch and the Juggernaut https://escholarship.org/content/qt2nr7t2zr/qt2nr7t2zr.pdf
44. Download book PDF - Springer Link https://link.springer.com/content/pdf/10.1007/978-94-011-5175-7.pdf
45. Computational Linguistics https://plato.stanford.edu/archives/fall2025/entries/computational-linguistics/
46. Proceedings of the 27th International Conference on ... https://aclanthology.org/volumes/C18-1/
47. Just Add Weights: Markov Logic for the Semantic Web https://dl.acm.org/doi/10.1007/978-3-540-89765-1_1
48. A Survey on Autonomous Scientific Discovery https://arxiv.org/html/2508.14111v2
49. Practices, opportunities and challenges in the fusion of ... https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2025.1590632/full

50. An Overview of Recent Advances in Natural Language ... https://www.mdpi.com/2076-3417/16/2/1122
51. (PDF) Ontological technologies for user modelling https://www.researchgate.net/publication/220094837_Ontological_technologies_for_user_modelling
52. Leveraging Large Language Models for Scalable and ... https://www.mdpi.com/2624-800X/5/3/55
53. Bayes-Entropy Collaborative Driven Agents for Research ... https://www.researchgate.net/publication/394293867_Bayes-Entropy_Collaborative_Driven_Agents_for_Research_Hypotheses_Generation_and_Optimization
54. The Unreasonable Effectiveness of Entropy Minimization in ... https://openreview.net/pdf?id=UfFTBEsLgI
55. S2CIE: semantic, syntactic, and context-based information ... https://www.sciencedirect.com/science/article/pii/S0160412025007378
56. Recent advances in biomedical literature mining https://academic.oup.com/bib/article/22/3/bbaa057/5838460
57. Combining hidden Markov models and latent semantic ... https://www.researchgate.net/publication/222580737_Combining_hidden_Markov_models_and_latent_semantic_analysis_for_topic_segmentation_and_labeling_Method_and_clinical_application
58. A Systematic Review of Topic Modeling Techniques for ... https://www.mdpi.com/2227-9032/14/2/282
59. Application and Prospects of Large Language Models in ... https://pmc.ncbi.nlm.nih.gov/articles/PMC12750412/
60. Knowledge-based Biomedical Data Science 2019 https://arxiv.org/pdf/1910.06710
61. Who thinks whom? It is the world which thinks us https://unesdoc.unesco.org/ark:/48223/pf0000121899
62. “So what if ChatGPT wrote it?” Multidisciplinary ... https://hal.science/hal-04586883/file/62927.VOR.pdf
63. Entropy, Volume 28, Issue 1 (January 2026) – 121 articles https://www.mdpi.com/1099-4300/28/1
64. Contributions to Explainable Artificial Intelligence https://cea.hal.science/tel-04615709v1/file/HDR_JPPoli_public.pdf
65. Entropy | Editor's choice Articles https://www.mdpi.com/journal/entropy/editors_choice
66. Entropy, Volume 26, Issue 10 (October 2024) – 77 articles https://www.mdpi.com/1099-4300/26/10

67. Inferring actual treatment pathways from patient records https://www.sciencedirect.com/science/article/pii/S1532046423002757
68. Keyphrase Identification Using Minimal Labeled Data with ... https://pmc.ncbi.nlm.nih.gov/articles/PMC10246160/
69. Exploring the impact of artificial intelligence-enhanced ... https://www.nature.com/articles/s41599-025-06033-x
70. A Comprehensive Bibliometric Analysis of Emerging ... https://ieeexplore.ieee.org/iel8/6287639/10820123/11080381.pdf
71. Computational Approaches to Predict Protein–Protein ... https://pubs.acs.org/doi/10.1021/acs.chemrev.3c00550
72. Findings of the Association for Computational Linguistics https://aclanthology.org/volumes/2024.findings-acl/
73. The challenge of uncertainty quantification of large ... https://arxiv.org/html/2504.05278v1
74. Early clonal extinction in glioblastoma progression ... https://www.sciencedirect.com/science/article/pii/S1535610823002416
75. Computer Science https://arxiv.org/list/cs/new
76. A Multi-Modal Graph Neural Network Framework for ... https://pmc.ncbi.nlm.nih.gov/articles/PMC12072649/
77. ACNP 62nd Annual Meeting: Poster Abstracts P251 – P500 https://www.nature.com/articles/s41386-023-01756-4
78. Findings of the Association for Computational Linguistics ... https://aclanthology.org/events/findings-2024/
79. A Survey on Large Language Models with some Insights ... https://arxiv.org/html/2501.04040v1
80. A Guide to Machine Learning Epistemic Ignorance, Hidden ... https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.70038
81. Research Frontiers in Machine Learning & Knowledge ... https://www.mdpi.com/2504-4990/8/1/6
82. Body of Knowledge - CS2023 https://csed.acm.org/wp-content/uploads/2024/01/Body-of-Knowledge-v1-bookmarksv2.pdf
83. Beyond mimicry: a framework for evaluating genuine ... https://pmc.ncbi.nlm.nih.gov/articles/PMC12832969/
84. Arxiv今日论文 | 2026-01-12 http://lonepatient.top/2026/01/12/arxiv_papers_2026-01-12

85. Findings of the Association for Computational Linguistics https://aclanthology.org/volumes/2024.findings-emnlp/
86. Partnership for the Assessment of Risks from Chemicals (PARC) https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e516799341&appId=PPGMS
87. Capabilities of Gemini Models in Medicine https://arxiv.org/html/2404.18416v2
88. Long Papers https://aclanthology.org/volumes/2024.acl-long/
89. Hw3 Stats Google 1gram | PDF | Internet Forum | Software https://www.scribd.com/document/917428277/Hw3-Stats-Google-1gram
90. Google Books Common Words | PDF | Experience https://www.scribd.com/document/852044092/Google-Books-Common-Words