From Narrative to Formalism: A Case Study in the Origin of Molecular Translation System

Anonymous Authors[1]

Abstract

We extend autoformalization into domain of natural sciences by exposing the logical structure of scientific narratives. We outline the Expert-in-the-Loop workflow that extracts axiom-like statements from the prose, uses LLM to generate Lean code enumerating propositions consistent with these axioms, and then identifies and interprets the most promising hypotheses. We demonstrate this approach in the investigation of the origin of translation as the key phase in the origin of life: we formalize an earlier proposed exaptation hypothesis, factorize an implicit signaling hypothesis, and derive a novel "signaling-first" hypothesis of the origin of translation.

1. Introduction

Formal logical structure forms the skeleton of scientific narratives far beyond mathematics. Therefore, the task of casting mathematical theories from natural-language descriptions into automatically verifiable form, aka *autoformalization* (Szegedy, 2020; Wu et al., 2022; Weng et al., 2025) has a rich counterpart that includes conversion of general scientific narratives into formal systems amenable to verification and automated exploration (cf. (Li et al., 2025; Liu et al., 2025).

Extraction of the logical skeleton of narratives and its translation into mechanized proofs is critical for rigor, scalability, and reproducibility of STEM research (cf. (Anonymous, 2024; Brunello et al., 2024)). We focus here on the Origin of Life (OoL) field as one of the cases where diverse subfields exist, each of them carries implicit formal structure in narratives, but these structures remain fragmented, difficult to reconcile, and without a path for systematic evaluate because they lack a common formal language.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

Let's consider a foundational problem in OoL, the origin of molecular translation, responsible for the existence of genetic code. A comprehensive treatment is offered in (Wolf & Koonin, 2007), where the authors put forward the exaptation hypothesis, H_0, arguing that translation developed when amino-acid binding enhanced catalytic activity of ribozymes in RNA world (Pressman et al., 2015) and was later repurposed for enzyme synthesis.

This raises a natural question: **Given the premises behind H_0, what alternative hypotheses one might construct?** Such a question applies to any scientific domain that lacks the formal rigor of mathematical theorems yet demands sound, auditable hypothesis generation.

In response, we sketch of a workflow that (1) parses scientific prose to extract axiom-like premises, (2) casts them as Lean (Moura & Ullrich, 2021) definitions, and (3) in Lean, enumerates the universe of all formal propositions consistent with those premises. The universe both recovers the propositions embedded in the original narrative and surfaces new, previously unstated hypotheses, including a novel "signaling-first" model for the origin of molecular translation.

The following exposition pursues two goals: (1) to execute a minimal end-to-end process in order to expose the key bottlenecks for future development, (2) to demonstrate that even the minimal instantiation of the process yields nontrivial, valuable insights.

2. Methodology

We frame the process as a cooperative game between a subject matter expert (SME) and a large language model (LLM) agent (Fig. 1A). Unlike in regular autoformalization (Lu et al., 2025a;b; Zhang et al., 2024), the logical backbone of the scientific prose is implicit. Expert-in-the-loop (EITL) mechanics help handling several fundamental challenges arising from this fact: (1) SME anchors the scope and prevents LLM from drifting away from the target task; (2) SME carries out quality assurance and domain alignment via audit of the axioms, propositions, and proofs; (3) SME assists LLM with implicit logic resolution and disambiguates both narrative-to-formalism and formalism-to-narrative conver-

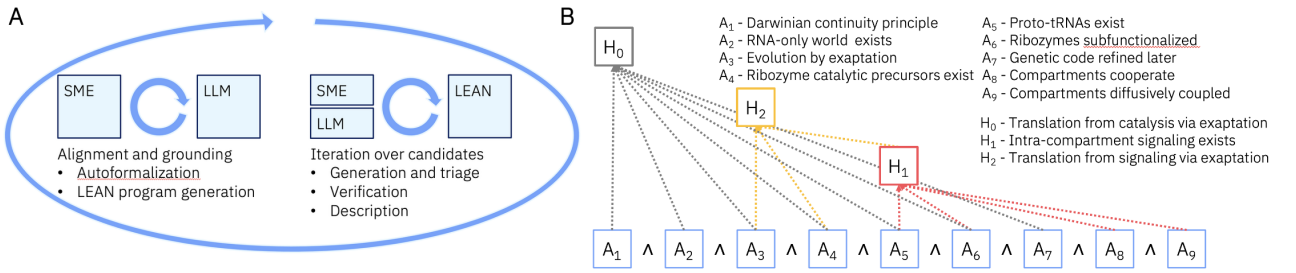

Figure 1. **Panel A**: General workflow of narrative conversion to formalism. The process is structured as a cooperative "game" between an SME and an LLM where Lean serves as the formal verification environment. The SME prevents drift of the scope of work, refines the types and definitions, carries out triage and prioritization of hypotheses, and helps to describe new hypotheses in natural language. The LLM generates Lean code including axiom type definitions and functions for the enumeration of the propositions derivable from the axioms. LLM does not generate new propositions - this is done by Lean program. **Panel B**: Formal representation of the original hypothesis about the origin of translation in the context of the origin of life and its hierarchical refactoring. A_i in blue boxes represent statements of the original paper that play the role of axioms. The original hypothesis H_0 is a conjunction of seven axioms $A_1 - A_7$ (gray arrows). Hypothesis about origin of signaling H_1 is refactored as a conjunction of four axioms $A_5, A_6, A_7, and A_8$ (red arrows). The new hypothesis that connects origin of translation to the origin of signaling is a conjunction of H_1, A_3 and A_4 (orange arrows).

sions; (4) SME determines the granularity of type definitions and decides if/when to introduce parameters, quantifiers, or higher-order structures; (5) SME performs hypotheses triage and navigates the universe of the generated propositions to isolate the most promising candidates.

We use Llama-3.3-70B-instruct (AI@Meta, 2024) as LLM agent. The LLM is allowed to independently extract the axiom-like statements from the original paper (Wolf & Koonin, 2007) and also performs assessment of the axioms extracted by SME for congruence between the two sets. LLM generates Lean code defining the agreed upon set of axioms and implements a combinatorial procedure in Lean for constructing propositions from the axioms. Validity of the propositions is verified in Lean. The universe of propositions is then searched to identify the items that offer alternatives to the original hypothesis H_0 where LLM generates a narrative counterpart of the formal statement and SME evaluates its novelty and relevance.

3. Results and Discussion

In the minimal version of the process, the LLM and SME agree on the set of 9 atomic premises $A_1 - A_9$ (see Fig. 1B), each declared in the sort *Prop* (universe of logical propositions) in Lean . The original hypothesis H_0 is a conjunction of 7 out of 9 axioms, $A_1 - A_7$ (Fig 1B) which is mechanically recovered and verified. We enumerate all non-empty conjunctions of the available axioms, arriving at 511 new propositions, each closed by a trivial *And.intro* proof.

Assessment of the propositions by SME and LLM leads to the isolation of a four-axiom core H_1 - a conjunction of axioms A_5, A_6, A_8, and A_9 described as "origin of translation". Finally, by adjoining two more premises (conjunction of H_1, A_3, and A_4) we arrive at a novel six-axiom hypothesis about origin of translation H_2, described as "origin of translation via exaptation from RNA-mediated intra-compartmental signaling system". Overall, H_0 depends on 7 axioms, H_1 on 4, and H_2 on six, which demonstrates both parsimony and modular reuse.

4. Conclusion

Our minimal EITL pipeline for narrative-to-formalism conversion reveals three core tracks for the future development:

- Bridging narrative and formalism: the bi-directional gap can be closed through cooperative, multi-agent workflows with expert oversight.
- Axiom-primitive design: it is non-trivial to decide how "rich" the primitives should be, from simple assertions, to quantified statements, to fully parametrized structures. It demands a systematically improvable process in multi-agent setting.
- Beyond conjunctions: enumeration of simple $\wedge$-chains surfaces trivial proofs, but general scientific arguments require richer constructs. It is necessary to extend the pipeline to generate and verify more complex propositions.

Even the minimal implementation was sufficient to reconstruct the original exaptation hypothesis H_0; factor out a four-axiom signaling core H_1; and derive a novel "signaling-first" hypothesis H_2 about the origin of translation in OoL.

References

AI@Meta. Llama 3 model card. 2024. URL https://github.com/meta-llama/llama3/blob/main/MODEL_CARD.md.

Anonymous. NL2FOL: Translating natural language to first-order logic for logical fallacy detection. In *Submitted to ACL Rolling Review - April 2024*, 2024. URL https://openreview.net/forum?id=lBc3MXcGqZ. under review.

Brunello, A., Ferrarese, R., Geatti, L., Marzano, E., Montanari, A., and Saccomanno, N. Evaluating llms capabilities at natural language to logic translation: A preliminary investigation. In *OVERLAY*, pp. 103–110, 2024. URL https://ceur-ws.org/Vol-3904/paper13.pdf.

Li, Q., Li, J., Liu, T., Zeng, Y., Cheng, M., Huang, W., Liu, Q., and Li, J. LINA: An LLM-driven neuro-symbolic approach for faithful logical reasoning, 2025. URL https://openreview.net/forum?id=3BoCwZFRJX.

Liu, T., Xu, W., Huang, W., Zeng, Y., Wang, J., Wang, X., Yang, H., and Li, J. Logic-of-thought: Injecting logic into contexts for full reasoning in large language models. In Chiruzzo, L., Ritter, A., and Wang, L. (eds.), *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 10168–10185, Albuquerque, New Mexico, April 2025. Association for Computational Linguistics. ISBN 979-8-89176-189-6. doi: 10.18653/v1/2025.naacl-long.510. URL https://aclanthology.org/2025.naacl-long.510/.

Lu, J., Wan, Y., Huang, Y., Xiong, J., Liu, Z., and Guo, Z. Formalalign: Automated alignment evaluation for autoformalization. In *The Thirteenth International Conference on Learning Representations*, 2025a. URL https://openreview.net/forum?id=B5RrIFMqbe.

Lu, J., Wan, Y., Liu, Z., Huang, Y., Xiong, J., Chengwu, L., Shen, J., Jin, H., Zhang, J., Wang, H., Yang, Z., Tang, J., and Guo, Z. Process-driven autoformalization in lean 4, 2025b. URL https://openreview.net/forum?id=k8KsI84Ds7.

Moura, L. d. and Ullrich, S. The lean 4 theorem prover and programming language. In *Automated Deduction – CADE 28: 28th International Conference on Automated Deduction, Virtual Event, July 12–15, 2021, Proceedings*, pp. 625–635, Berlin, Heidelberg, 2021. Springer-Verlag. ISBN 978-3-030-79875-8. doi: 10.1007/978-3-030-79876-5_37. URL https://doi.org/10.1007/978-3-030-79876-5_37.

Pressman, A., Blanco, C., and Chen, I. The rna world as a model system to study the origin of life. *Current Biology*, 25(19):R953–R963, Oct 2015. ISSN 0960-9822. doi: 10.1016/j.cub.2015.06.016. URL https://doi.org/10.1016/j.cub.2015.06.016.

Szegedy, C. A promising path towards autoformalization and general artificial intelligence. In *Intelligent Computer Mathematics: 13th International Conference, CICM 2020, Bertinoro, Italy, July 26–31, 2020, Proceedings*, pp. 3–20, Berlin, Heidelberg, 2020. Springer-Verlag. ISBN 978-3-030-53517-9. doi: 10.1007/978-3-030-53518-6_1. URL https://doi.org/10.1007/978-3-030-53518-6_1.

Weng, K., Du, L., Li, S., Lu, W., Sun, H., Liu, H., and Zhang, T. Autoformalization in the era of large language models: A survey, 2025. URL https://arxiv.org/abs/2505.23486.

Wolf, Y. I. and Koonin, E. V. On the origin of the translation system and the genetic code in the rna world by means of natural selection, exaptation, and subfunctionalization. *Biology Direct*, 2(1):14, May 2007. ISSN 1745-6150. doi: 10.1186/1745-6150-2-14. URL https://doi.org/10.1186/1745-6150-2-14.

Wu, Y., Jiang, A. Q., Li, W., Rabe, M. N., Staats, C., Jamnik, M., and Szegedy, C. Autoformalization with large language models. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN 9781713871088.

Zhang, L., Quan, X., and Freitas, A. Consistent autoformalization for constructing mathematical libraries. In Al-Onaizan, Y., Bansal, M., and Chen, Y.-N. (eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 4020–4033, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.emnlp-main.233. URL https://aclanthology.org/2024.emnlp-main.233/.

A. Lean, axioms

```
axiom continuity : Prop
axiom continuity_axiom : continuity

axiom rna_world : Prop
axiom rna_world_axiom : rna_world

axiom exaptation : Prop
axiom exaptation_axiom : exaptation

axiom ribozyme_precursors : Prop
axiom ribozyme_precursors_axiom : ribozyme_precursors

axiom proto_tRNAs : Prop
axiom proto_tRNAs_axiom : proto_tRNAs

axiom subfunctionalization : Prop
axiom subfunctionalization_axiom : subfunctionalization

axiom code_refinement : Prop
axiom code_refinement_axiom : code_refinement

axiom compartments : Prop
axiom compartments_axiom : compartments

axiom diffusive_exchange : Prop
axiom diffusive_exchange_axiom : diffusive_exchange
```

B. Lean, enumeration of candidate propositions, conjunction only

```
@[simp]
def conjList : List Prop → Prop
| []      => True
| p :: ps => p ∧ conjList ps

def candidate_props : List Prop :=
  allAxioms
    |>.subsets
    |>.filter (fun xs => xs ≠ [])
    |>.map (fun xs => conjList (xs.map axiomProp))

#eval candidate_props.length
```

C. Lean, signaling proposition and trivial proof from axioms by And.intro

```
def signaling : Prop :=
  proto_tRNAs ∧ subfunctionalization ∧ compartments ∧ diffusive_exchange

theorem signaling_from_axioms
  (h2 : proto_tRNAs)
  (h6 : subfunctionalization)
  (h8 : compartments)
  (h9 : diffusive_exchange) :
  signaling := by
  -- Expand `signaling` to see its nested-∧ structure:
  unfold signaling
  -- Now the goal is `proto_tRNAs ∧ (subfunctionalization ∧ (compartments ∧
    diffusive_exchange))`.
  -- First split off `proto_tRNAs`:
  apply And.intro
  · exact h2
```

```
  -- The remaining goal is `subfunctionalization ∧ (compartments ∧ diffusive_exchange)`.
  -- Split that as well:
  apply And.intro
  · exact h6
  -- Now the remaining goal is `compartments ∧ diffusive_exchange`.
  apply And.intro
  · exact h8
  · exact h9
```