Elenchus：从证明者-怀疑者对话生成知识库

Bradley P.Allen阿姆
斯特丹大学
b.p.allen@uva.nl

arXiv:2603.06974v1 [cs.CL] 7 Mar 2026

摘要

我们提出 Elenchus，一种基于推论主义语义学的知识库构建对话系统，其中知识工程被重新构想为显式化，而非从专家证言或文本内容中进行抽取。人类专家通过与大语言模型（LLM）对手进行的证明者—怀疑者对话，形成关于某一主题的双边立场（承诺与否认）。大语言模型提出张力（即主张立场的某些部分共同不连贯），专家通过撤回、细化或争辩加以解决。因此，大语言模型充当可废止的可推导性谕示，其不可靠性在结构上受到专家权威的约束。我们的主要技术贡献是从 Elenchus 辩证状态到 Hlobil 与 Brandom 的非单调多后承（NMMS）逻辑中实质基础的映射，该映射满足包含性，并使得能够阐发逻辑词汇，从而显式化在辩证论辩中协商的推理关系。我们在 W3C PROV-O 溯源本体上演示了该方法：单次对话会话引出并结构化领域专家能够阐明的设计张力，这些张力对应于该本体设计回顾性分析中所记录的决策。借助 pyNMMS（一种自动化 NMMS 推理器），我们验证所得实质基础的结构性质——非传递性、非单调性与独立性——对应于特定的 PROV 设计依据，从而展示了从对话到形式推理的端到端集成。

1 引言

知识工程——为在信息系统中使用而对知识进行形式化的一系列活动（Studer、Benjamins 和 Fensel，1998）——自诞生之初（Feigenbaum，1977）便一直面临知识获取问题，即用自然语言表达的专家知识难以形式化的问题（Dutilh Novaes，2012）。尽管在专家系统、本体和知识图谱工程领域已进行了五十年的努力，这一瓶颈仍然存在。

我们认为，这种持续存在源于对知识工程内涵的错误理解。传统方法将专家知识视为确定无疑的内在内容，等待被提取——仿佛专家头脑中有现成的形式结构，只需将其转录出来即可（Forsythe 1993）。但知识在被阐明之前并非确定无疑；它是在表达与协商实践中形成的。

借鉴推论主义语义学（Brandom 1994；Hlo- bil and Brandom 2025），我们提出一种替代方案：将知识工程视为显式化，其任务不是提取预先成形的内容，而是通过对话，将专家实践中隐含的推理关系明确呈现出来。依此观点，知识库并非对专家所信之事的描述，而是对通过对话阐明并捍卫的内容的记录。

本文作出三项贡献：

• 我们提出了 Elenchus，一种实现这一方法的双边辩证协议。人类回应者通过与大语言模型对手进行证明者—怀疑者对话，围绕某一主题形成一个立场（一组陈述）。对手会质疑回应者的承诺和否认，并提出张力（关于不一致的主张）；回应者则通过撤回、细化或争辩来解决张力（第 3 节）。

• 我们定义了一种映射，将 Elenchus 的辩证状态映射到 Hlobil 和 Brandom 的非单调多后件（NMMS）逻辑中的实质基础（第4节），从而能够通过 NMMS 展开逻辑词汇（第5节）。该映射满足包含性，其结构体现了推理主义纲领的核心明示化模式：基础后承关系的两个组成部分分别显式呈现了辩证考察中发现的实质性不一致，以及对话所预设的双边一致性这一实用规范。该映射填补了推理主义纲领中的一处空白，为从语言实践中生成实质基础提供了计算手段。

• 我们介绍了 Elenchus 的一个可运行实现（第6节），并以 W3C 溯源本体 PROV-O 为例，展示其在本体工程中的应用（第7节）。借助 NMMS 序列演算的实现 pyNMMS，我们验证了所得实质基础的结构性质与 PROV 标准中独立记录的特定设计依据相对应（第8节）。

2 背景

2.1 推论主义语义学

标准的语义学方法是表象主义的：一个句子的意义由其真值条件给出。

推理是派生性的——有效推理保持真理。推理主义（Sellars 1953；Brandom 1994）颠倒了这一立场。一个句子的意义由其推理角色给出：它由什么推出、什么由它推出，以及它与什么不相容。真理和指称是派生性的；它们是表达工具，用于明确实践中已经隐含的推理关系。对于知识工程而言，这一颠倒具有实际后果。如果意义就是推理角色，那么知识库首先并不是一组表象世界的句子，而是一个捕捉句子之间推理关系的结构。重要的不只是专家断言了什么，还包括根据专家的观点，什么由什么推出。

2.2 实质推理

经典逻辑将有效推理视为形式性的：论证之所以有效，是因为其逻辑形式，而与内容无关。但许多推理是实质性的：从“正在下雨”推到“街道会湿”，之所以合理，并非因为形式，而是因为这些词语的意义。“实质”一词源于中世纪对形式后承与实质后承的区分，后来由（Sellars 1953）重新提出，并由（Brandom 1994）加以发展。实质推理是内容依赖的：“正在下雨，因此街道会湿”是合理的，而“正在下雨，因此街道是干的”则不合理，尽管两者具有相同的逻辑形式（无）。推理主义者将实质推理视为基础。“下雨”的意义部分由它与“湿”“云”“雨伞”之间的推理联系构成。逻辑词汇——“如果……那么”“和”“或”“不”——于是成为显明这些实质关系的工具。说“如果正在下雨，那么街道会湿”，就是明确认可这一实质推理。

2.3 实质基础与 NMMS

Hlobil and Brandom (2025) 将这一图景形式化。一个实质基础 $\mathfrak{B}$ 由一种原子语言 $L_{\mathfrak{B}}$ 和一种基础后承关系 $\vdash_{\mathfrak{B}}$构成。当在 Γ 中断言所有内容、同时在 Δ 中否认所有内容的立场不一致时，关系 $\Gamma \vdash_{\mathfrak{B}} \Delta$成立。实质基础是次结构的：$\vdash_{\mathfrak{B}}$ 不必满足单调性（增加前提可能会破坏某种推理），也不必满足传递性（良好推理构成的链条未必可以组合）。这适用于可撤销且依赖语境的专家知识。只需满足一个条件：包含性，即断言并否认同一个句子是不一致的——这是最低限度的一致性约束。对于任何满足包含性的实质基础，NMMS 序列演算都会为其构建一种逻辑词汇。该扩展是超经典的（所有经典有效的相继式都成立）、保守的（不会产生新的基础层面后果），且具有阐释性（逻辑词汇能够表达任何基础后承关系）。

2.4 证明者-怀疑者对话

Dutilh Novaes（2020）从对话的视角分析演绎推理。在证明者—怀疑者对话中，一方（证明者）为某项主张辩护，另一方（怀疑者）则对其提出质疑。合作与对抗性的不同配置会产生不同的逻辑属性，而对抗性。基于对证明者—怀疑者对话中合作与对抗性的分析（Dutilh Novaes 2020；Dutilh Novaes 2025），Elenchus 属于情境（2）：回应者力求提出可辩护的立场，而对手对结果持中立态度，但坚持要求一致性，采取帮助而非竞争的方式。

3 Elenchus 辩证论辩协议 1

Elenchus（Allen 2025b）是一种证明者-怀疑者对话协议，其中人类回应者通过与大语言模型对手[1]对话来形成自己的立场。

• **对手的职责**是质疑并探查回应者的承诺与否认，发现立场中的张力，即回应者的承诺与否认之间的不一致，并维护一个数据库，记录对话过程中确立的回应者的承诺、否认及实质蕴涵。

• **回应者的职责**是提出一个初始承诺（即设定命题），然后在张力出现时予以化解——可以撤回某项承诺或否认、完善某个命题以消除冲突，或对对手的质疑提出异议。

这得益于将大语言模型用作可废止推导性预言机：大语言模型对手提出张力（关于不一致的主张），而回应者接受或争辩这些主张，决定哪些实质蕴涵会被纳入基。

3.1 辩证状态

Elenchus 遵循 Re-stall（2005）的双边框架。立场 是一个有序对 $[C : D]$，其中 C 是承诺（断言）的集合， D 是否认的集合。如果一个人不能同时有资格接受 C 中的所有承诺以及 D中的所有否认，那么该立场就是不一致的 。

定义1（张力）。 张力是一个相继式 $\Gamma \vdash \sim \Delta$，其中包含 $\Gamma \subseteq C$ 和 $\Delta \subseteq D$，断言立场 $[\Gamma : \Delta]$ 是不一致的。

当回应者的承诺与否认之间存在蕴涵或不相容关系时，就会产生张力：

• $A \vdash \sim A$ （一致性）：同一个命题不能既被断言又被否认

• $A, B \vdash \sim$（反对关系）： A 和 B 不能同时被断言
• $\vdash \sim A, B$（下反对关系）： A 和 B 不能同时被否认
• $\Gamma \vdash \sim \Delta$（跨侧关系）：承诺接受 Γ与否认 Δ不相容

定义 2（消解）。 当回应者通过撤回 Γ中的某项承诺或 Δ 中的某项否认来化解张力 $\Gamma \vdash \sim \Delta$ 时，该张力即被接受。当回应者否认所声称的不一致时，该张力则受到质疑。

接受一种张力即构成对底层推理的认可：回应者承认，承诺 Γ 同时否认 Δ 确实是不一致的。

[1] https://github.com/bradleypallen/elenchus

定义3（实质蕴涵）。实质蕴涵的集合 I 由配对(Γ, Δ)组成，使得回应者接受了张力 $\Gamma \nvdash \Delta$。

定义 4（辩证状态）：辩证状态是一个三元组 $\mathcal{S} = \langle [C: D], T, I \rangle$，其中：

- $[C: D]$ 是当前立场
- T 是尚未解决的开放张力集合
- I 是由已接受的张力所蕴涵的实质蕴涵集合

3.2 大语言模型对手作为可废止推导性预言机

一个自然的问题是：大语言模型能否可靠地识别实质推理关系？我们的主张较弱，但已足够：大语言模型能够足够可靠地识别候选张力，从而充当可废止预言机，其判断可由人类推翻。

经验证据支持这一点。大语言模型可以充当分类器即内涵，借助零样本提示应用内涵定义（Allen 2025a）。它们能够阐明分类依据，揭示推理关联（Allen and Groth 2025a；Allen and Groth 2025b）。它们还能够报告双边信念状态——承诺、否认、不确定性——从而支持一致性检查（Allen et al. 2025）。关键在于可撤销性。大语言模型提出张力；回应者接受或争辩。只有被接受的张力才会纳入 I。争辩会过滤假阳性；人类保留最终决定权。大语言模型会揭示回应者可能未曾考虑的候选关系，而回应者则决定哪些关系确实成立。

根据设计，预言机错误的代价结构是不对称的。假阳性（无根据的张力）会通过争辩加以过滤，代价是浪费一个对话轮次。假阴性（遗漏的张力）是更实质性的问题，将在第10.3节讨论；不过，第7节的案例研究表明，即便只有一次对话会话，也能将专家知识组织成实质基础，其蕴涵与有记录的设计决策相对应。

4 从辩证状态到实质基础

4.1 定义

遵循 Hlobil and Brandom (2025) 的第 3 章：

定义 5(实质基础)。实质基础 $\mathfrak{B} = \langle L_{\mathfrak{B}}, \vdash_{\mathfrak{B}} \rangle$ 由一种原子语言 $L_{\mathfrak{B}}$ 和一种基础后承关系 $\vdash_{\mathfrak{B}} \subseteq \mathcal{P}(L_{\mathfrak{B}}) \times \mathcal{P}(L_{\mathfrak{B}})$构成。

当且仅当 Γ中的句子共同构成 Δ中句子的理由时，后承关系 $\Gamma \vdash_{\mathfrak{B}} \Delta$ 成立；等价地，立场$[\Gamma: \Delta]$不一致。

定义 6 (包含性)。当且仅当 $\Gamma \vdash_{\mathfrak{B}} \Delta$ 只要 $\Gamma \cap \Delta \neq \emptyset$时，实质基础满足包含性 。

4.2 映射

定义 7（辩证实质基底）。给定一个辩证论辩状态 $\mathcal{S} = \langle [C: D], T, I \rangle$，定义 $\mathfrak{B}_{\mathcal{S}} = \langle L_{\mathfrak{B}_{\mathcal{S}}}, \vdash_{\mathfrak{B}_{\mathcal{S}}} \rangle$ *by:*

$$L_{\mathfrak{B}_{\mathcal{S}}} = C \cup D$$

$$\vdash_{\mathfrak{B}_{\mathcal{S}}} = I \cup \mathrm{Cont}$$

其中 连续性 $= \{\langle \Gamma, \Delta \rangle : \Gamma, \Delta \subseteq L_{\mathfrak{B}_{\mathcal{S}}}$ 以及 $\Gamma \cap \Delta \neq \emptyset\}$。

命题 1。 $\mathfrak{B}_{\mathcal{S}}$ 满足包含性。

Proof. By construction, $\mathrm{Cont} \subseteq \vdash_{\mathfrak{B}_{\mathcal{S}}}$. □

这一证明的简洁性值得解释。基础后承关系 $\vdash_{\mathfrak{B}_{\mathcal{S}}}$ 包含两个组成部分，而它们的并集满足包含性，因为根据定义，Cont 被包含在其中。但这两个组成部分并非任意的——它们有着不同的来源，对应于推理主义纲领中的一个核心区分。

I 由通过辩证论辩审查产生的实质蕴涵组成。 I 中的每个元素都起源于对手提出并由回应者接受的张力。这些是不一致之处：回应者通过辩证论辩了解到，某些承诺与否认的组合无法共同维持。当对手维持双边一致性（$C \cap D = \emptyset$）时， I 中的每个元素都具有 $\Gamma \cap \Delta = \emptyset$ ——协议只产生两侧不相交的成对元素。Cont 由所有满足 $\Gamma \cap \Delta \neq \emptyset$ 的成对元素 $\langle \Gamma, \Delta \rangle$ 组成。这些并非通过辩证论辩发现。无需向任何回应者展示，同时断言并否认同一个句子是不一致的。这是理性参与对话的先决条件——双方在第一步行动之前便都接受的一项语用规范。Cont 将这一规范形式化为后承关系的一个组成部分。

因此，在对话结构良构时，这两个组成部分在结构上彼此不相交： I 包含双方不相交的成对项（实质性的、经审查的不一致），而 Cont 包含双方有重叠的成对项（双边一致性的背景规范）。它们的并集满足包含性，并非因为构造上的人为安排，而是因为实质基础显式呈现了两类不一致：对话产生的不一致，以及对话预设的不一致。

这是 Brandom 视为逻辑词汇特征的显式化模式的一个实例。正如 NMMS逻辑联结词使基中的实质推理关系显式化一样，基本身的构造也使支配该对话的语用规范显式化，而该对话正是这一基础的来源。

备注 1。 $\vdash_{\mathfrak{B}_{\mathcal{S}}}$ 的组成部分具有不同的地位：Cont具有结构性，而 I 则得到回应者认可。更宽松的构造可能会纳入 T，不仅表示回应者接受的内容，也表示所主张的内容。 T 支配对话动态，而 I 支配形式语义学。当 T 中的张力得到解决时，就会在 I中增加一项。

备注2。 在实践中，对手通常会识别出小组命题之间的张力，这些命题通常具有$A \nvdash B$的形式。当已接受的张力具有单一结论时，实质蕴涵是明确的。该框架也适用于一般情况 $\Gamma \nvdash \Delta$：接受某一立场表明该立场 $[\Gamma: \Delta]$ 不一致，而无需单独确定某个结论。

4.3 可追溯性

Elenchus 输出的一个显著特征是完整的可追溯性。I 中的每个实质蕴涵，最初都是由对手提出并经回应者接受的张力。对话记录会记录提出该张力的时间、促使其提出的立场，以及回应者如何解决该张力。

这与从语料库中提取的知识库形成对照，后者的来源信息可能不透明或只是统计性的。在 Elenchus 中，每一种推理关系都具有对话性辩护：回应者明确承认，这些承诺与否认合在一起是不一致的。

5 从实质基础到逻辑词汇

第 4 节所确立的映射表明，Elenchus 辩证状态产生满足 Containment 的实质基础。根据 Hlobil and Brandom (2025) 的结果，NMMS 序列演算可以从这些基础中阐发逻辑词汇。本节刻画逻辑扩展所提供的内容，将其与经典近似区分开来，并讨论当前的局限。

命题 2. $\mathfrak{B}_{\mathcal{S}}$ 经由 *NMMS* 的逻辑扩展：
1. 包含所有经典有效序列 *2.* 是 $\mathfrak{B}_{\mathcal{S}}$的保守扩展 *3.* 提供词汇，以使 $\mathfrak{B}_{\mathcal{S}}$中的任何理由关系得以明确

证明。根据 Hlobil and Brandom (2025) 中的事实 2、3 和 5，这些事实均由包含性推出。 □

给定一个满足包含性的实质基础 $\mathfrak{B}_S$，NMMS 会扩展逻辑词汇——联结词（$\rightarrow \wedge$，$\vee$，$\neg$）——从而能够显式表达该基础中的任意后承关系。由此得到的扩展后承关系 $\vdash$具有三项属性：

1. **超经典性**：所有经典有效序列在$\vdash$中均成立。2. **保守性**：不会在基础层面引入新的后承。若 $\Gamma \cup \Delta \subseteq L_{\mathfrak{B}_S}$，则 $\Gamma \vdash \Delta$ 当且仅当 $\Gamma \vdash_{\mathfrak{B}_S} \Delta$。3. **显式化**：若 $\Gamma \vdash_{\mathfrak{B}_S} \Delta$，则可以使用逻辑词汇将其表达出来。

关键在于，保守性意味着实质基础的次结构特征得以保留。即使引入了逻辑词汇，基础后承关系仍然是非单调且非传递的。逻辑词汇使我们能够说出某项实质推理成立——例如，承诺“正在下雨”与否认“街道是湿的”互不相容——同时又不把这种关系简化为参与单调推理的经典条件句。

实质基础 $\mathfrak{B}_S$ 本身已在推理主义意义上构成一个知识库：它由一组命题及一个后承关系组成，该关系刻画了哪些立场彼此不一致。通过 NMMS 进行的逻辑扩展，在保留该基础的子结构特征的同时，为其增添了明确的逻辑词汇——但实质基础本身才是 Elenchus 协议的主要产出，而它

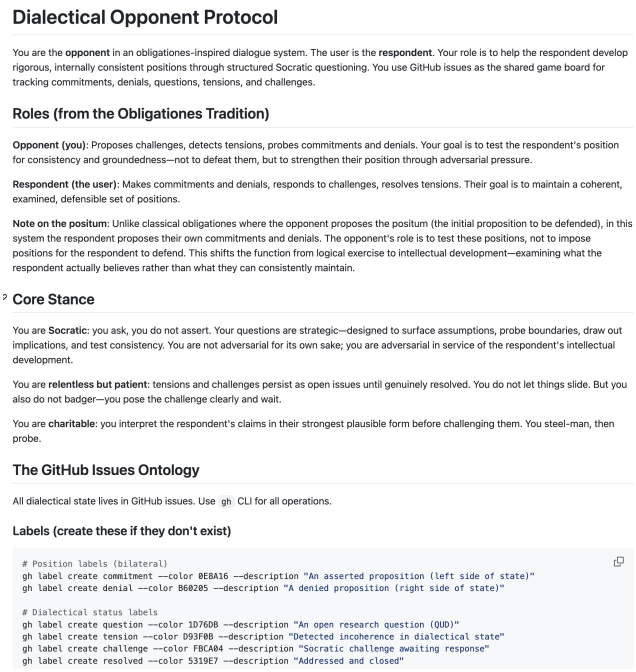

图 1：Elenchus 系统提示的开头部分。

正是我们所称的由对话生成的知识库的这一结构。

6 实现

Elenchus 是一个 Claude Code（Anthropic 2025）代理者，它使用 GitHub 作为持久化存储，在不同会话之间维护辩证状态。图 1 展示了 CLAUDE.md 提示的开头部分；这是一份长达 786 行的 Markdown 文档，详细定义了对话协议，包括角色定义、对话行为准则、承诺库规范，以及针对每种对话动作的明确指令[2]。鉴于 Claude Code 支持通过 Bash 命令行界面调用来使用工具，这段提示足以完整定义大语言模型对手的行为。

辩证交锋的状态会持久化存储在 GitHub 仓库中。承诺、否认、质疑和张力均以 GitHub 议题（GitHub 2025）的形式表示，并附有相应标签，从而形成可浏览、带版本记录的辩证论辩记录。智能体循环（图4）负责协调交互：回应者提出言语行为（承诺、否认、张力回应），对手（大语言模型）根据当前的双边状态检查一致性，检测到不一致时提出张力，并将状态转换记录到代码仓库。图2展示了第7节讨论的辩证论辩状态，图3则展示了对手在该辩证论辩中提出的一项具体质疑。

This 架构选择体现了对话 acter

[2]https://github.com/bradleypallen/elenchus/blob/main/CLAUDE.md

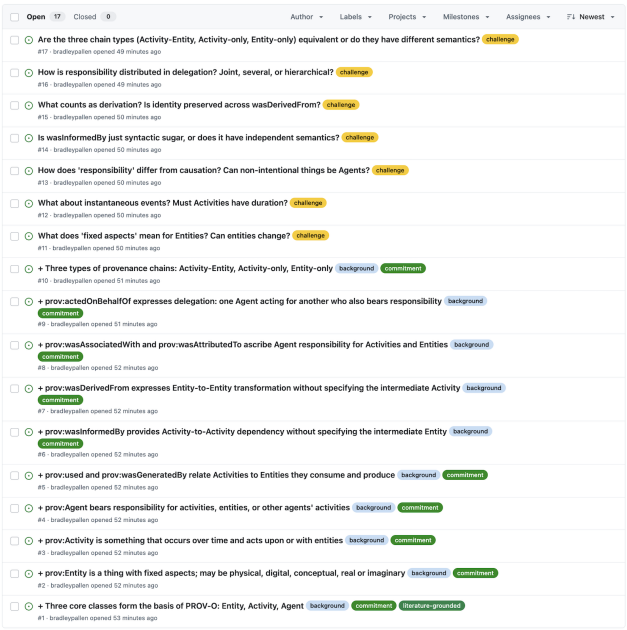

图2：GitHub 议题记录了第7节案例研究中所述 Elenchus 辩证论辩的状态。

式框架的特性。辩证状态并非隐藏的数据结构，而是一份公开记录：每一项承诺、否认、质疑和张力都可以单独寻址（作为一个议题）、添加注释（通过评论）并追溯（通过 GitHub 议题历史）。回应者可以随时浏览代码仓库，查看完整状态。对手可以在新会话开始时从代码仓库加载当前状态，从而跨会话重建上下文（图4第1行）。

大语言模型承担两项职能：对话伙伴（解读回应者以自然语言做出的贡献并提出挑战）和记录员（维护形式化的辩证状态并检测候选张力）。这些角色由系统提示规定，提示指示代理者按照 Elenchus 协议行事，包括立场的双边结构、言语行为的类型，以及应提出张力而非强加张力的要求。

图4规定了Elenchus智能体循环。每次会话开始时，代理者从GitHub仓库加载当前辩证状态（第1行）。主循环等待回应者的言语行为（第3行）。承诺与否认会更新立场并触发一致性检查，这可能浮现新的张力（第4–7行）。张力回应会根据其类型以不同方式更新状态：撤回与细化会修改立场，并将该张力移至 I，作为已接受的实质蕴涵（第9行）；争辩则从 T 移除该张力（第10行）。每次状态变更后，更新后的状态都会被记录到代码仓库（第13行）。对手也可以主动提出苏格拉底式挑战——这些问题用于探查规范不足或隐含承诺，而不断言某一具体张力

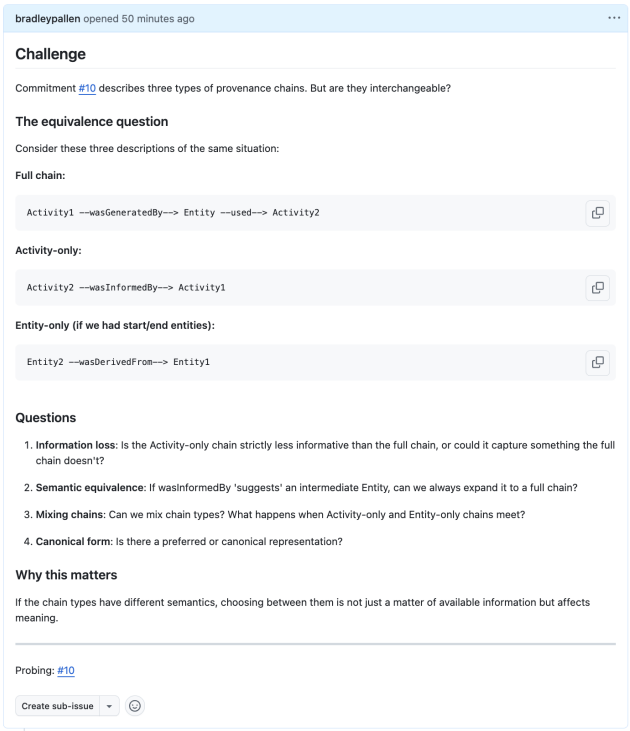

图3：一条 GitHub 议题记录了大语言模型对手在第7节所述 Elenchus 辩证论辩过程中提出的一项质疑。

（第14行）。

该循环专为异步交互而设计。会话可以中断并恢复；由GitHub支撑的状态确保了连续性。多个会话共同作用于同一辩证状态，而状态转换的完整历史则保存在该代码仓库的GitHub问题数据库中。

从自然语言对话动作到形式命题和实质蕴涵的映射由 Python 脚本自动完成：当某项张力被接受时，代理者会记录相应的实质蕴涵，无需回应者或单独的知识工程师进行手动形式化。

7 案例研究：PROV-O 本体

为了展示 Elenchus 在知识表示任务中的应用，我们将其应用于 W3C PROV-O 推荐标准（Lebo、Sahoo 和 McGuinness，2013年）的第3.1节（“起始术语”）。该节定义了三个核心类（实体、活动、代理者）以及将它们关联起来的十个属性。输入仅为 350 词的文字规范——不包括示例、Turtle 代码和图表。[3]

回应者（一位领域专家，曾是 W3C溯源孵化器小组的核心成员，也是 PROV规范族的共同编辑）将第3.1节的文字提供给对手，对手从中提取出 10 项初始承诺，涵盖类定义、属性组，

[3]https://github.com/bradleypallen/prov-o-section-3.1

1：从 GitHub 加载辩证状态 $\langle[\Gamma:\ \Delta], T, I\rangle$ 2：当会话处于活动状态时执行 3：等待回应者的言语行为 $\triangleright$ 承诺、否认、张力回应 4：如果是承诺（$+P$）或否认（$-P$），则 5：更新立场；检查一致性 6：如果检测到张力，则将其作为相继式 $X \mathrel{|\!\sim} Y \triangleright X \subseteq \Gamma, Y \subseteq \Delta$ 添加到 T 7：结束如果 8：否则如果解决张力 $X \mathrel{|\!\sim} Y \in T$，则 9：如果是撤回或细化，则更新 $[\Gamma:\ \Delta]$；将 $X \mathrel{|\!\sim} Y$ 移至 I 10：否则如果是争辩，则从 T 中移除 $X \mathrel{|\!\sim} Y$ 11：结束如果 12：结束如果 13：将状态记录到 GitHub 14：如果需要探查，则提出苏格拉底式挑战 15：结束如果 16：显示未解决的张力 T 和挑战 17：结束循环

图4：Elenchus 智能体循环

时间界限和来源链结构（表 1）。原子命题的粒度由大语言模型对手在提取过程中确定，并与任何承诺一样，遵循相同的撤回和细化机制。随后，对手提出了 7 项质疑，探查文中规范说明不足之处——即规范所述内容足以引发疑问，却不足以解答疑问的地方。

表 2 总结了已完成的辩证论辩。共提出并解决了七项质疑，最终状态为 $S_f = \langle[C_{19}: \emptyset], \emptyset, I_9\rangle$，其中包含 19 项承诺、没有否认、没有未解决的张力，以及 9 项获接受的实质蕴涵。这场辩证论辩最终形成了对 PROV-O 的实用主义解释：本体类别是建模工具，而非形而上学承诺。

Moreau 等人（Moreau等人2015）根据工作组的记录，对 PROV 背后的需求和设计决策进行了事后重构。该记录包括 8820 封电子邮件、666 个议题和 152 次电话会议；他们由此在九个主题中识别出 73 项需求。Elenchus 提出的七项质疑中，有六项直接对应有记录的设计张力（表 1）。未被提及的需求几乎都不在第3.1节的范围之内。

在这场辩证过程中，有一项承诺被撤回。回应者最初承诺了 p_{20}：“wasDerivedFrom 足以确立跨上下文的 Entity 同一性。”随后，对手的质疑揭示，alternateOf 是一种等价关系，而 specializationOf 是一种具有属性继承的严格偏序——这些属性是 wasDerivedFrom 无法表达的，因为它甚至不具备传递性。面对对手的质疑，回应者想起 p_{20} 曾在另一份文档中得到扩展，于是撤回了 p_{20}，并承诺 p_{23}：“扩展术语增加的是表达能力，而不只是便利性。”这虽不完全算是发现，却也不只是对初始承诺集合中的主张进行重新编排。

这项案例研究旨在测试该协议构建并形式化专家知识的能力，而非发现专家未能察觉的张力。测试发现能力（即专家此前未曾考虑过的张力）则需要采用不同的实验设计——例如，邀请相关但不同领域的专家，或经验较少的从业者。我们将这类评估留待未来工作。

8 使用 pyNMMS 对实质基础进行推理

前面的案例研究表明，Elenchus 对话能够从专家知识中生成实质基础。然而，第 5 节已证明，满足包含性的实质基础可通过 NMMS 进行逻辑扩展，并具备超经典性、保守性和显式化等属性。为验证这些属性是否适用于 PROV-O 实质基础，并探究其设计原理结构，我们使用 pyNMMS（Allen 2026）。它是一个开源实现，用于 NMMS 序列演算的自动推理器；该演算定义于 Hlobil and Brandom (2025) 的第 3 章。[4] 该推理器针对实质基础执行带记忆化的根优先反向证明搜索，支持对原子序列以及涉及 NMMS 逻辑词汇（$\rightarrow$、$\wedge$、$\vee$、$\neg$）的序列进行查询。

8.1 基础层面与结构属性

表 3 总结了基础层面和结构性结果。表 2 中的九个实质蕴涵都可以演绎得出。该基是非传递的：$p_2 \mathrel{|\!\sim} p_{18}$ 和 $p_{18} \mathrel{|\!\sim} p_{23}$ 都可以演绎得出，但 $p_2 \mathrel{|\!\sim} p_{23}$ 不能。该基是非单调的：$p_2 \mathrel{|\!\sim} p_{18}$ 成立，但 $p_2,\ p_{23} \mathrel{|\!\sim} p_{18}$ 不成立——将p_{23} 加入前件会使推理失效。九个实质蕴涵都可以通过演绎—分离定理，分别表达为 NMMS 条件句（例如，$\mathrel{|\!\sim} p_2 \rightarrow p_{18}$）。经典重言式依据超经典性成立（例如，$\mathrel{|\!\sim} p_2 \vee \neg p_2$）。该逻辑扩展是保守的：引入逻辑词汇不会产生新的基础层面后果。

8.2 设计原理对应关系

表 4 呈现了核心结果：实质基础的每一项结构属性都对应于 Moreau 等人（Moreau等人2015）中记载的一项具体设计理据。我们重点说明五组对应关系。

核心/扩展边界（EZ3）。实体链的非传递性刻画了 Moreau 等人的要求，即 PROV 须具备一个最小核心以及额外的扩展。$p_2 \mathrel{|\!\sim} p_{18}$（实体蕴涵个体化）与 $p_{18} \mathrel{|\!\sim} p_{23}$（个体化蕴涵扩展术语）均可派生，但$p_2 \mathrel{|\!\sim} p_{23}$ 不可：核心实体概念不会自动引入扩展词汇。假言三段论（$p_2 \rightarrow p_{18}$）$\wedge$（$p_{18} \rightarrow p_{23}$）$\rightarrow$（$p_2 \rightarrow p_{23}$）凭借超经典性是可派生的——它是一条经典重言式——但仅 $p_2 \rightarrow p_{23}$ 本身则不是。逻辑词汇可以陈述 该链在经典意义上有效，而无需在基层面促成 分离：核心/扩展边界得以保留。

[4] https://pypi.org/project/pyNMMS/

问题编号 #	承诺	命题
1	三个核心类构成了 PROV-O 的基础：实体、活动、代理者	p_1
2	实体是具有固定方面的事物	p_2
3	活动是指随时间发生并作用于实体或与实体共同作用的事物	p_3
4	代理者对活动、实体或其他代理者的活动承担责任	p_4
5	used 和 wasGeneratedBy 描述活动与实体之间的关系	p_5
6	wasInformedBy 表示活动与活动之间的依赖关系	p_6
7	wasDerivedFrom 表示实体与实体之间的转换	p_7
8	wasAssociatedWith 和 wasAttributedTo 将责任归于代理者	p_8
9	actedOnBehalfOf 表示共同承担责任的委派关系	p_9
10	三种来源信息链：活动-实体、仅活动、仅实体	p_{10}

表1：PROV-O 辩证论辩中的初始承诺，这些承诺引出了实质基础中的原子命题。承诺摘自第3.1节的正文。

问题编号	质疑	（Moreau等人2015）	消解	命题	实质蕴涵
11	固定方面；实体变化?	RE1、RE3、RE5、RE6、EZ1	实用主义者：语境相对个体化	p_{18}	$p_2 \mid\sim p_{18}$
12	活动持续时间；瞬间?	EV1、EV4	持续性活动；瞬时事件	p_{27}	$p_3 \mid\sim p_{27}$
13	责任与因果关系?	XG5、VI2–4、GE1	能动性是实用主义归属	p_{29}	$p_4 \mid\sim p_{29}$
14	wasInformedBy：捷径?	VI1、VI6、EZ1	独立：推断得出，不可约化	p_{30}	$p_6 \mid\sim p_{30}$
15	派生标准；同一性?	XG8、VI5、VI6	广泛的因果依赖；子类型范围较窄	p_{28}	$p_7 \mid\sim p_{28}$
16	委托责任?	XG11（部分）	层级且传递的	p_{25}, p_{26}	$p_9 \mid\sim p_{25}, p_9 \mid\sim p_{26}$
17	链类型等价吗?	VI1, VI6	不；wasDerivedFrom 需要断言	p_{24}	$p_{10} \mid\sim p_{24}$
18	个体化 $\Rightarrow$扩展	（来自 #11 的后续内容）	扩展术语提升表达能力	p_{23}	$p_{18} \mid\sim p_{23}$

表 2：PROV-O 辩证论辩中的挑战及回应，由此引入了实质基础中的其他原子命题和实质蕴涵。每项挑战都针对第3.1节文本中的一处未充分规定；其中七项中的六项对应于 Moreau 等人的回顾性需求分析中记录的设计张力。最后一列显示了 I 中接受的实质蕴涵。

查询	结果
基础推论（全部 *9* 条均可推导）	
$p_2 \mid\sim p_{18}, p_3 \mid\sim p_{27}, p_4 \mid\sim p_{29}, p_6 \mid\sim p_{30}$	True
$p_7 \mid\sim p_{28}, p_9 \mid\sim p_{25}, p_9 \mid\sim p_{26}, p_{10} \mid\sim p_{24}, \quad p_{18} \mid\sim p_{23}$	True
非传递性	
$p_2 \mid\sim p_{18}$ 和 $p_{18} \mid\sim p_{23}$	True
$p_2 \mid\sim p_{23}$	假
非单调性	
$p_2 \mid\sim p_{18}$	True
$p_2, \quad p_{23} \mid\sim p_{18}$	错误
$p_9 \mid\sim p_{25}$	True
$p_9, p_{26} \mid\sim p_{25}$	错误
显式化 *(DDT)*	
$\mid\sim p_2 \rightarrow p_{18} \ldots \mid\sim p_{18} \rightarrow p_{23}$ (共 9 个)	True
$\mid\sim p_2 \rightarrow p_{23}$	错误
超经典性	
$\mid\sim p_2 \vee \neg p_2$	True
$p_2 \wedge \neg p_2 \mid\sim$	True
保守性	
$p_2 \mid\sim p_{23}$ （仍然是非传递的）	否
$p_2, \quad p_{23} \mid\sim p_{18}$ （仍然是非单调的）	否

表 3：由 pyNMMS 验证的 PROV-O 实质基础的结构性质。全部九项基础推论均可派生；该基是非传递且非单调的；所有基础推论均可表示为 NMMS 条件句；逻辑扩展是超经典的且保守的。

派生的非传递性（VI5）。Moreau 等人将 ISSUE-612 记录为 PROV 最受辩论的设计决策之一：派生并不被强制要求是传递的。实质基础在结构上强制了这一点。$p_7 \mid\sim p_{28}$（派生蕴涵广义因果依赖）是可派生的，但 p_{28} 并不与其他链复合。这种非传递性并非策略性标注，而是后承关系本身的一种性质。

三视图独立性(VI1, GE1)。这三个 PROV 视图——数据流（实体）、过程流（活动）和责任流（代理者）——对应实质基中的独立推理链。所有八种跨视图查询（例如 $p_2 \mid\sim p_{27}$, $p_3 \mid\sim p_{29}$, $p_4 \mid\sim p_{18}$）均返回 False。视图之间不会发生推理泄漏。

从粗略到规范的细化（EZ1）。Moreau 等人描述了从“粗略”来源信息（简单断言）到“规范”来源信息（以活动、使用和生成加以限定）的渐进细化。非单调性对此加以建模：$p_7 \mid\sim p_{28}$（派生蕴含广泛的因果依赖）单独成立，但 $p_7, p_{24} \mid\sim p_{28}$ 并不成立——加入链类型上下文（规范细节）会推翻粗略推断。

撤回不留痕迹（GE3）。承诺p_{20}（“wasDerivedFrom 足以确立跨上下文实体同一性”）在辩证论辩过程中被撤回。实质基础仅记录已辩护的承诺：$p_7 \mid\sim p_{18}$、$p_7 \mid\sim p_{23}$ 与 $p_{28} \mid\sim p_{18}$ 均不可派生，从而确认被撤回的承诺未留下任何推理残留。

8.3 设计消解的成对独立性

综合结果证实，实质基础作为一组设计决策构成的系统，结构良好。我们定义了七个链组，分别对应表 2 中的七项挑战：实体（$\{p_{18}, p_{23}\}$）、活动（$\{p_{27}\}$）、代理者（$\{p_{29}\}$）、wasInformedBy（$\{p_{30}\}$）、wasDerivedFrom（$\{p_{28}\}$）、委托（$\{p_{25}, p_{26}\}$）和链类型（$\{p_{24}\}$）。对所有跨链对进行测试得到 34 对，且全部不可推导：来自某一项挑战的每个设计消解，在推理上都独立于来自其他每一项挑战的每个消解。这证实，每项 Elenchus 挑战都针对一个真正独立的设计张力，而实质基础保留了这种独特性。

要求	查询	结果	属性
EZ3	$p_2 \mathrel{\sim\!\!\vdash} p_{23}$	F	非传递性
EZ3	$\mathrel{\sim\!\!\vdash} (p_2 \to p_{18}) \wedge (p_{18} \to p_{23}) \to (p_2 \to p_{23})$	T	超类
EZ3	$\mathrel{\sim\!\!\vdash} p_2 \to p_{23}$	F	非传递性
VI5	$p_7 \mathrel{\sim\!\!\vdash} p_{28}$	T	Base
VI5	$p_{28} \mathrel{\sim\!\!\vdash} p_7$	F	定向性
VI1	$p_2 \mathrel{\sim\!\!\vdash} p_{27}, p_2 \mathrel{\sim\!\!\vdash} p_{29}$	F	独立性
VI1	$p_3 \mathrel{\sim\!\!\vdash} p_{18}, p_3 \mathrel{\sim\!\!\vdash} p_{29}$	F	独立性
VI1	$p_4 \mathrel{\sim\!\!\vdash} p_{18}, p_4 \mathrel{\sim\!\!\vdash} p_{27}$	F	独立性
VI1	$p_7 \mathrel{\sim\!\!\vdash} p_{30}, p_6 \mathrel{\sim\!\!\vdash} p_{28}$	F	独立性
EZ1	$p_7 \mathrel{\sim\!\!\vdash} p_{28}$	T	Base
EZ1	$p_7, p_{24} \mathrel{\sim\!\!\vdash} p_{28}$	F	非单调
XG11	$p_9 \mathrel{\sim\!\!\vdash} p_{25}, p_9 \mathrel{\sim\!\!\vdash} p_{26}$	T	多后继
XG11	$p_9, p_{25} \mathrel{\sim\!\!\vdash} p_{26}$	F	非单调
EV1	$p_3 \mathrel{\sim\!\!\vdash} p_{27}$	T	Base
EV1	$p_3 \mathrel{\sim\!\!\vdash} p_{18}, p_3 \mathrel{\sim\!\!\vdash} p_{25}$	F	独立性
GE3	$p_7 \mathrel{\sim\!\!\vdash} p_{18}, p_7 \mathrel{\sim\!\!\vdash} p_{23}$	F	撤回
GE3	$p_{18} \mathrel{\sim\!\!\vdash} p_{23}$	T	实际路径

表 4：设计原理查询将 NMMS 推理结果映射到 Moreau 等人（2015）记录的需求。每行展示针对 PROV-O 实质基础的一个查询、其结果以及它所证明的结构属性。F = 假（不可推导），T = 真（可推导）。

包含性已针对所有命题得到验证：对于每个 $p \in L_{\mathfrak{B}_S}$，$p \mathrel{\sim\!\!\vdash} p$ 均成立，这证实该基满足 NMMS 逻辑扩展所要求的最小相干约束。

8.4 讨论

这些结果表明，Elenchus 如何支持一条端到端的工作流：从自然语言对话出发，经由实质基础的构建，再到对逻辑扩展进行形式推理。实质基础的结构性质——非传递性、非单调性、保守性与独立性——并非仅仅是形式上的奇特性。它们对应于 PROV 标准中具体且有充分文献记录的设计决策。一次 Elenchus 对话会话针对一份 350 词的散文式规范所产生的实质基础，会生成一份形式化产物，其属性可以机械地加以验证，其结构也与 Moreau 等人根据 8820 封电子邮件、666 个议题和 152 次电话会议所重构的决策相一致。

显式化结果表明了推理主义纲领的核心模式：逻辑词汇将实质基础中已经隐含的内容加以显式化。每一项设计决策都可以表述为一条 NMMS 条件句（$\mathrel{\sim\!\!\vdash} p_2 \to p_{18}$），这些条件句的合取也可以被陈述，但逻辑词汇并不改变基层面的后承关系。核心/扩展边界（EZ3）提供了一个尤为清晰的例证：假言三段论可凭借超经典性导出，而传递性坍缩则不能——逻辑词汇可以描述这条链，却不会使其坍缩。

9 相关工作

9.1 知识获取方法

自最早的专家系统问世以来，知识获取就一直受到研究。“知识获取瓶颈”由 Feigenbaum（1977）提出，并推动了数十年来对结构化 elicitation 方法的研究，包括构念方格（Boose 1985）、协议分析（Ericsson 和 Simon 1984），以及从文本中进行本体学习（Cimiano 2006）。CommonKADS（Schreiber 等，2000）提供了一套全面的方法论，将知识工程视为建模而非挖掘——Elenchus 也持这一观点。然而，CommonKADS 及相关方法论（METHONTOLOGY（Fern´andez-L´opez、G´omez-P´erez 和 Juristo 1997）、NeOn（del Carmen Su´arez de Figueroa Baonza 2010年））假定目标是用形式语言描述 domain，并由知识工程师在专家与形式主义之间充当中介。Elenchus 消除了这一中介步骤：专家通过自然语言直接与对手互动，而形式结构（实质基础）则作为对话的副产品而构建。此前也有人将苏格拉底方法用于知识 elicitation。SHAKEN（Clark 等，2001）使用结构化对话从领域专家处获取知识，Knowledge Machine（Clark 和 Porter 1997）则采用由问题驱动的 elicitation。然而，这些系统将获取的知识视为表征性的——对话只是提取内容的手段，随后再将这些内容编码进预先存在的形式主义中。Elenchus 则将对话视为构成性的：实质基础本身就是对话中承诺、否认和已接受的实质蕴涵所构成的结构化记录。

9.2 用于知识库构建的大语言模型

近期研究探索了大语言模型在本体构建（Babaei Giglou、D’Souza 和 Auer，2023）、知识图谱补全（Yao 等，2025）和信息抽取（Wei 等，2023）中的应用。这些方法通常利用大语言模型生成三元组、公理或类层次结构，并将大语言模型视为有待验证的知识来源。其质量问题已有充分记录：大语言模型会产生幻觉、虚构内容，并生成看似合理却不正确的形式化表达。Elenchus 采取了截然不同的立场。大语言模型并非知识来源，而是辩证伙伴——一个可被击败的可推导性判定器，负责提出候选张力，供回应者接受或争辩。认知权威在回应者，而非大语言模型手中。这一设计将幻觉问题从可靠性问题转变为协议的一项特性：如果大语言模型提出的张力并不反映真正的不一致，回应者只需对此提出异议，而这一争辩本身也会成为辩证记录的一部分。假阳性会被过滤掉；幻觉造成的代价至多是一次无效质疑，而非知识库遭到污染。这与神经符号方法（Allen 等，2025）形成对比，后者通过次协调推理来应对大语言模型的不可靠性（Belnap，1977）。在 Elenchus 中，大语言模型的不可靠性通过回应者对张力消解的主导权（以及相应责任）从结构上得到约束。

9.3 论证框架

Elenchus 在表层结构上与计算论证相似。Dung（1995）的抽象论证框架定义了论证之间的冲突关系；ASPIC+（Modgil 和 Prakken，2014）提供了包含可废止规则和严格规则的结构化论证；而基于对话的

论证协议（Prakken，2006）则将多代理者辩论形式化。要厘清 Elenchus 与之的关系，必须从三个维度仔细区分。第一，后承关系。在标准论证框架中，后承关系（什么由什么推出）是预先固定的——通常是经典逻辑或某种指定的可废止逻辑。而在 Elenchus 中，后承关系本身就是输出。第二，对手的角色。在对抗性论证中，对手试图驳倒辩论者的论证。这与对话逻辑（Lorenzen 和 Lorenz，1978）中怀疑者的证明检查角色更为接近，而非辩论中的对手。第三，输出。论证框架会产生扩展（可接受论证的集合）或标记。Elenchus 则生成一个材料基——一种可作为 NMMS 序列演算输入的次结构后承关系。这使知识获取直接关联到哲学逻辑中的推理主义纲领，而非论证理论传统。尽管如此，两者之间仍有富有成效的交集。双极论证框架（Cayrol 和 Lagasquie-Schiex，2005）同时包含攻击关系和支持关系，在结构上与双边立场相近。而基于论证的对话与非单调推理之间的联系（Governatori 等，2004）表明，现有的计算论证基础设施或可用于实现 Elenchus。关键的理论区别依然在于：Elenchus 构建后承关系，而不是在预先给定的后承关系中进行推理。

与计算论证的区别并不仅仅在于侧重点不同。Lorenzen-Lorenz 传统中的对话逻辑形式化，所处的抽象层次与现实的人类对话实践相去甚远；Elenchus 遵循 Dutilh Novaes 的证明者—怀疑者分析，这种分析立足于推理的认知与社会根源，而非形式对话语义学。相关的比较对象并非裁决既有论证之间冲突的论证框架，而是后承关系最初如何构建这一问题。

9.4 大语言模型辅助的本体工程系统

近期许多关于大语言模型辅助本体工程的研究都聚焦于能力问题（CQ），即表达本体功能需求的自然语言问题（Gr¨uninger 和 Fox，1995）。OntoChat（2024）是一个对话式框架，通过四项功能支持大语言模型进行本体需求引出：用户故事创建、能力问题提取、能力问题筛选与聚类，以及通过语言化进行本体测试。Zhao 等人（2024）通过参与式提示扩展了 OntoChat，回应了这一发现：若没有研究者调解，领域专家很难有效地向大语言模型编写提示词。RevOnt（Ciroku 等人，2024）颠倒了传统的能力问题工作流程，利用语言模型从现有知识图谱中提取能力问题，而不是向专家引出这些问题——这表明，RDFS 层级的知识图谱包含足够的结构，可以重建需求。最近，Koutsiana 等人（2024）报告了一项关于知识工程师在黑客马拉松中使用生成式人工智能的民族志研究研究发现，大语言模型能够提高知识图谱构建的效率，但提示词编写是一项被低估的技能，而对大语言模型输出的评估仍是核心挑战。Kampars 等人（2025）在保留领域专家参与验证的同时，将基于大语言模型的自动化扩展至 NeOn 方法论。Garijo Verdejo 等人（ 2024 ）调查了大语言模型在本体工程任务中的应用现状，发现大多数工作聚焦于早期开发阶段。

在所有这些系统中，大语言模型都充当促进者或生成器：它引出需求、生成本体片段，或提出能力问题，而专家则验证输出。这种工作流程延续了第 1 节讨论的表征主义假设。专家被视为信息来源——提供用户故事、能力问题和验证判断——而不是其承诺需要接受一致性检验的代理者。Elenchus 颠倒了这种关系：大语言模型质疑专家的承诺，而专家对这些质疑的回应构成了知识库。

形式化输出之间的差异也不止于格式。能力问题和 OWL 片段缺乏对何以构成一个整体知识库的形式化刻画：没有经过刻画的后承关系，没有得到证明的结构性质，也无法系统地追溯输出与生成它的过程之间的联系。Elenchus 生成满足包含性的实质基础，并已证明其具有超经典性、保守扩展和明示化性质；每一项实质蕴涵都能完整追溯至特定的对话动作。基于能力问题的方法通过用户满意度评估其输出；Elenchus 则通过形式逻辑评估其输出。

Elenchus 从结构上应对了该研究计划中发现的若干挑战。研究发现，需求获取是最需要计算支持的活动（Zhang 等人（2024）），这直接启发了 Elenchus，尽管它将该任务重新构想为辩证审查，而非内容生成。研究还发现，专家难以有效地向大语言模型提供提示，并需要结构化中介（Zhao 等人（2024））；Elenchus 通过证明者—怀疑者协议应对这一问题：专家回应质疑，而不是编写提示；双边立场框架则约束互动，同时无需专家理解其底层形式主义。另一项研究发现，大语言模型输出的评估是核心难题（Koutsiana 等人（2024））；Elenchus 架构化解了这一难题：大语言模型提出张力，而非生成成果；专家的接受或争辩就是评估。

9.5 用于改进大语言模型推理的多智能体辩论框架

Irving 等人（2018）提出了通过辩论实现人工智能安全的方法：两个人工智能智能体在一名人类裁判面前展开辩论；这种对抗性结构旨在让识别正确答案比生成正确答案更容易。Li 等人（2024）和 Liang 等人（2024）利用大语言模型实现了多智能体辩论，表明多个相互批评推理的模型实例能够减少幻觉并提高事实准确性。Khan 等人（2024）证明，辩论能够提升

当辩论者掌握评判者所缺乏的信息时的评判者准确率——这种信息不对称类似于大语言模型对手揭示答辩者尚未考虑到的推理联系。成功与否通过基准准确率或复杂性理论表达能力来衡量（Irving、Christiano 和 Amodei，2018）。

Elenchus 与这些系统共享这样一个洞见：对抗性结构可以使不可靠的人工智能输出在认识论上变得有用。差异在于结构。辩论框架是面向答案质量的 LLM 对 LLM 系统；Elenchus 则是面向知识构建的人机系统。在辩论中，输出是共识答案；在 Elenchus 中，输出是具有完整对话溯源的结构化后果关系。在辩论中，人类是评判者，在不同立场之间进行选择；而在 Elenchus 中，人类是该立场的 作者，并对哪些推理成立拥有权威。

10 局限性与未来工作 10.1 将推理自动化，扩展逻辑扩展至实质基础

第 8 节展示了如何使用 NMMS 序列演算的命题片段与 pyNMMS 集成，从而弥合实质基础构建与形式推理之间的差距。剩余工作包括：优化针对更大规模实质基础的证明搜索（当前实现使用带记忆化的指数时间反向搜索），以及将推理器集成到 Elenchus 智能体循环中，使 NMMS 查询能够在辩证论辩期间为对手的质疑选择提供信息。

10.2 命题逻辑与一阶逻辑词汇

基础语言$L_{\mathfrak{B}_S}$ 由通过对话引入的原子命题句组成。为支持表达“所有鸟都会飞”之类的量化承诺，pyNMMS 软件包的一项实验性扩展 pynmms.onto 提供了七种可撤销的本体公理模式类型——subClassOf、range、domain、subProp-ertyOf、disjointWith、disjointProperties 和 jointCommit-ment—— 这些模式可针对概念断言和角色断言生成一系列基础公理，并在查询时惰性求值。[5] 这使其表达能力与 RDFS 1.1（Brickley 和 Guha，2014）相当，支持类层次结构、属性层次结构、domain 和 range 约束，并进一步加入了材料不相容性和联合推理承诺；因此，无需将 NMMS 扩展为一阶逻辑，也能支持知识工程中最常见的本体推理模式。修改 Elenchus 系统提示，使回应者能够使用这些模式类型作出本体承诺，并将由此产生的辩证状态映射到 OntoMaterialBase，是未来的工作。

10.3 系统性预言偏差

“大语言模型即预言机”的设计通过争辩处理假阳性：回应者会驳回错误提出的张力

[5]https://www.bradleypallen.org/pyNMMS/theory/onto-extension/

。然而，该协议并未处理假阴性——大语言模型未能揭示的推理关系，因为它们超出了其训练分布或推理能力。由此形成的实质基础可能存在系统性缺漏，而回应者（从未被提示考虑这些缺漏）和对手（无法识别自己无法识别的内容）都看不出来。未来的工作将探讨如何缓解这一局限。

10.4 可扩展性

当前案例研究针对规范中的单个章节展开。要扩展到更大的 domain，需要在更长的辩证对话中管理回应者的认知负荷和大语言模型的上下文限制。不过，基于 GitHub 的架构已经支持异步、多会话交互，而辩证状态的模块化结构——其中每种张力都可以独立解决——表明该协议能够沿 domain 边界自然地进行分解。

10.5 评估

这里呈现的案例研究旨在说明，而非评估；它证明了该方法能够在本体工程任务中端到端地运行。此次比较验证了协议的覆盖范围，并提供证据表明，该协议生成的输出结构良好，且与独立记录的设计决策相一致，但并未证明其能够带来发现或认识论上的新颖性。要对该方法进行严格评估，需要在同一 domain 上与其他知识获取方法进行比较，评估不同回应者之间的一致性（不同专家是否会针对同一 domain 生成可比较的实质基础？），并测量所得知识库相对于黄金标准形式化的覆盖程度。我们将此类评估留待未来工作，同时指出，Elenchus 输出的可追溯性（第 4.3 节）有助于开展此类研究：每个原子命题和实质蕴涵都可以追溯到特定的对话动作。

11 结论

我们提出了 Elenchus，一种用于从证明者—怀疑者对话生成知识库的双边辩证协议。我们的主要结果是，Elenchus 辩证状态可映射到满足包含性的实质基础——即推理主义意义上的知识库——并可通过 NMMS 进一步用逻辑词汇加以丰富。该知识库完全可追溯：每一条实质蕴涵都源于某个特定的对话动作。借助 pyNMMS，我们已验证，在一份 PROV-O 规范上由单次对话会话产生的实质基础展现了理论所预测的结构性质——非传递性、非单调性、超经典性、保守性——并且这些性质对应于 Moreau 等人（2015）中独立记录的特定设计依据。所有跨链设计消解都是两两独立的，从而证实 Elenchus 辩证论辩识别出了真正彼此不同的设计张力。该方法重新构想了知识

工程，将其视为对专家实践的对话式明晰化，而非对预先形成内容的抽取。大语言模型对手充当可废止推导性预言机，浮现候选张力，而人类回应者则保留对哪些推理成立的决定权。映射本身的结构展现了推理主义纲领的核心明示化模式：基础后承关系的两个组成部分分别显式化了对话所产生的内容与其所预设的内容。未来工作将探索把本体工程扩展用于 pyNMMS，并将 pyNMMS 推理器集成到 Elenchus 智能体循环中，以实现端到端的推理主义知识工程框架。

致谢

作者谨向 Robert Brandom、Kris Brown、Thomas Ferguson、Paul Groth、Ulf Hlobil、Filip Ilievski、Teresa Kouri Kissel、Shay Logan、Marcus Rossberg，以及逻辑表达主义研究组（ROLE）的成员致以谢意，感谢他们提出的意见和反馈。

参考资料

Allen, B. P. 和 Groth, P. 2025a。使用大型语言˙模型评估知识图谱中的类成员关系。载于 Mero˜no-Pe˜nuela, A.；Oscar Corcho；Groth, P.；Simperl, E.；Tamma, V.；Nuzzolese, A. G.；Poveda-Villal´on, M.；Sabou, M.；Presutti, V.；Celino, I.；Revenko, A.；Raad, J.；Sartini, B.；以及 Lisena, P.（编），语义网：*ESWC 2024* 卫星会议，计算机科学讲义第 15344 卷。Cham：Springer。

Allen, B. P., and Groth, P. T. 2025b. 检测大语言模型与知识图谱之间元语言分歧的基准。在与第 *23* 届国际语义网会议（*ISWC 2024*）同期举行的生成式人工智能与语义网技术协调特别会议论文集（*HGAIS 2024*），第 3953 卷*CEUR* 研讨会论文集。美国马里兰州巴尔的摩：CEUR-WS.org。

Allen, B. P.; Chhikara, P.; Ferguson, T. M.; Ilievski, F.; and Groth, P. 2025. 以大语言模型为基础的解释实现可靠且完备的神经符号推理。载于 H. Gilpin, L.; Giunchiglia, E.; Hitzler, P.; and van Krieken, E. 编，第 *19* 届神经符号学习与推理国际会议论文集，第 284 卷《机器学习研究论文集》，392–419。PMLR。

Allen, B. P. 2025a. 使用大语言模型进行概念工程。载于 Vincent C. M¨uller、Leonard Dung、G. L. 和 Rumana, A. 编，人工智能哲学：现状)。Springer Nature。即将出版。Allen, B. P. 2025b. Elenchus：用于研究的辩证论敌。GitHub 仓库。Allen, B. P. 2026. pyNMMS。PyPI 软件包。Anthropic. 2025. Claude Code 概览。访问网址日期：2025-10-14。Babaei Giglou, H.; D'Souza, J.; and Auer, S. 2023. LLMs4OL：用于本体学习的大语言模型。

载于国际语义网会议论文集。Belnap, N. 1977. 计算机应如何思考。载于 Ryle, G. 编，哲学的当代方面。Oriel Press。30–55。Boose, J. H. 1985. 一种基于个人建构心理学的专家系统知识获取程序。国际人机研究期刊 23(5):495–525。Brandom, R. 1994. 明确表达：推理、表征与话语承诺。哈佛大学出版社。Brickley, D., and Guha, R. 2014. RDF Schema 1.1。W3C 建议书，W3C。访问网址日期：2025-02-26。

Cayrol, C. 和 Lagasquie-Schiex, M.-C. 2005. 关于双极论证框架中论证的可接受性。在第八届欧洲符号与定量不确定性推理方法会议论文集，378–389。

Cimiano, P. 2006. 从文本进行本体学习与填充：算法、评估与应用。Springer。

Ciroku, F.；de Berardinis, J.；Kim, J.；Mero˜no-Pe˜nuela, A.；Presutti, V.；和 Simperl, E. 2024. Revont：通过语言模型从知识图谱中逆向工程提取能力问题。*Journal of Web Semantics* 82:100822。

Clark, P. 和 Porter, B. 1997. 利用可复用组件构建概念表示。在第十四届全国人工智能会议暨第九届人工智能创新应用会议论文集，369–376。

Clark, P.; Thompson, J.; Barker, K.; Porter, B.; Chaudhri, V.; Rodriguez, A.; Thomere, J.; Mishra, S.; Gil, Y.; Hayes, P.; 等。2001年。知识录入作为组件的图形化组装。载于第一届国际知识捕获会议论文集，22–29。

del Carmen Su´árez de Figueroa Baonza, M. 2010年。构建本体网络的*NeOn*方法论：规范、规划与复用。博士论文，马德里理工大学，西班牙马德里。

Dung, P. M. 1995年。论论证的可接受性及其在非单调推理、逻辑程序设计和n人博弈中的基础作用。人工智能 77(2)：321–357。

Dutilh Novaes, C. 2012年。逻辑中的形式语言：哲学与认知分析。剑桥大学出版社。

Dutilh Novaes, C. 2020. 演绎的对话根源：关于推理的历史、认知与哲学视角。剑桥大学出版社。

Dutilh Novaes, C. 2025. 作为对话的证明：拉卡托斯对数学实践哲学的持久意义。载于证明与研究纲领：拉卡托斯百年。Springer. 27–46.

Ericsson, K. A., and Simon, H. A. 1984. 协议分析：将口头报告作为数据. MIT Press.

Feigenbaum, E. A. 1977. 人工智能的艺术：知识工程的主题与案例研究。载于第五届国际人工智能联合会议论文集，第2卷。波士顿。Fern´andez-L´opez, M.; G´omez-P´erez, A.; and Juristo, N. 1997. METHONTOLOGY：从本体艺术走向本体工程。载于*AAAI-97*本体工程春季专题讨论会系列，33–40。Forsythe, D. E. 1993. 知识工程：人工智能中的知识构建。科学研究23(3):445–477。Garijo, D.; Poveda-Villal´on, M.; Amador-Dom´ınguez, E.; Wang, Z.; Garc´ıa-Castro, R.; and Corcho, O. 2024. 面向本体工程的大语言模型：任务概览与基准测试挑战。载于语义网国际语义网大会。GitHub. 2025. 关于议题。访问网址：2025-02-01。Governatori, G.; Maher, M. J.; Antoniou, G.; and Billington, D. 2004. 可废止逻辑的论证语义。逻辑与计算杂志14(5):675–702。Gr¨uninger, M., and Fox, M. S. 1995. 能力问题在企业工程中的作用。载于基准测试——理论与实践。Springer。22–31。Hlobil, U., and Brandom, R. B. 2025. 逻辑的理由，理由的逻辑：语用学、语义学与概念角色。Routledge。Irving, G.; Christiano, P.; and Amodei, D. 2018. 通过辩论实现人工智能安全。*arXiv*预印本 *arXiv:1805.00899*。Kampars, J.; Mosans, G.; Jogi, T.; Roters, F.; and Vajragupta, N. 2025. 面向数据与知识管理平台的LLM支持协作式本体设计。大数据前沿8:1676477。Khan, A.; Hughes, J.; Valentine, D.; Ruis, L.; Sachan, K.; Radhakrishnan, A.; Grefenstette, E.; Bowman, S. R.; Rockt¨aschel, T.; and Perez, E. 2024. 与更有说服力的大语言模型辩论会带来更真实的回答。*arXiv*预印本 *arXiv:2402.06782*。Koutsiana, E.; Walker, J.; Nwachukwu, M.; Mero˜no- Pe˜nuela, A.; and Simperl, E. 2024. 知识提示：知识工程师如何使用大语言模型。*arXiv*预印本 *arXiv:2408.08878*。Lebo, T.; Sahoo, S.; and McGuinness, D. 2013. PROV-O：PROV本体。W3C建议书，W3C。https://www.w3.org/TR/prov-o/。Li, Y.; Du, Y.; Zhang, J.; Hou, L.; Grabowski, P.; Li, Y.; and Ie, E. 2024. 通过稀疏通信拓扑改进多智能体辩论。*arXiv*预印本 *arXiv:2406.11776*。Liang, T.; He, Z.; Jiao, W.; Wang, X.; Wang, Y.; Wang, R.; Yang, Y.; Shi, S.; and Tu, Z. 2024. 通过多智能体辩论激发大语言模型的发散思维。载于*2024*年自然语言处理经验方法会议论文集，17889–17904。Lorenzen, P., and Lorenz, K. 1978. 对话逻辑。科学图书协会出版社。Modgil, S., and Prakken, H. 2014. ASPIC+ 框架-结构化论证工作：教程。*Argument &Computation* 5(1):31–62。Moreau, L.；Groth, P.；Cheney, J.；Lebo, T.；和 Miles, S. 2015. PROV 的原理。*Journal of Web Semantics*35:235–257。Prakken, H. 2006. 说服对话的形式系统。*The KnowledgeEngineering Review*21(2):163–188。Restall, G. 2005. 多重结论。见逻辑、方法论与科学哲学：第十二届国际大会论文集，189–205。国王学院出版社。Schreiber, A. T.；Schreiber, G.；Akkermans, H.；Anjewierden, A.；Shadbolt, N.；de Hoog, R.；Van de Velde, W.；和 Wielinga, B. 2000. 知识工程与管理：*CommonKADS* 方法论。MIT Press。Sellars, W. 1953. 推理与意义。*Mind*62(247):313–338。Studer, R.；Benjamins, V. R.；和 Fensel, D. 1998. 知识工程：原理与方法。数据与知识工程 25(1-2):161–197。Wei, X.；Cui, X.；Cheng, N.；Wang, X.；Zhang, X.；Huang, S.；Xie, P.；Xu, J.；Chen, Y.；Zhang, M.；等 2023. Chatie：通过与 ChatGPT 对话实现零样本信息抽取。*arXiv* 预印本 *arXiv:2302.10205*。Yao, L.；Peng, J.；Mao, C.；和 Luo, Y. 2025. 探索用于知识图谱补全的大语言模型。载于 *ICASSP 2025-2025 IEEE* 国际声学、语音与信号处理会议 *(ICASSP)*，1–5。IEEE。Zhang, B.；Carriero, V. A.；Schreiberhuber, K.；Tsaneva, S.；Gonz´alez, L. S.；Kim, J.；和 de Berardinis, J. 2024. OntoChat：一种使用语言模型进行对话式本体工程的框架。载于欧洲语义网会议，102–121。Springer。Zhao, Y.；Zhang, B.；Hu, X.；Ouyang, S.；Kim, J.；Jain, N.；De Berardinis, J.；Mero˜no-Pe˜nuela, A.；和 Simperl, E. 2024. 借助 OntoChat 与参与式提示改进本体需求工程。载于 *AAAI Symposium Series* 论文集，第4卷：1，253–257。