

Rule-Guided Reasoning via Multi-Agent Test-Driven Program Synthesis: A Neuro-Symbolic Approach

Author Name

Affiliation

email@example.com

Abstract

Large Language Models (LLMs) often struggle with complex rule-guided reasoning tasks due to challenges in rule recall, logic consistency, and precise numerical computation. To address these limitations, we propose a novel multi-agent framework that transforms natural language rules into executable and verifiable program segments. Our approach employs a rule-decomposition mechanism to break down complex instructions into atomic, manageable sub-rules. We then introduce a collaborative multi-agent system comprising Code Generation, Test Case Generation, and Verification agents. By adopting a test-driven development (TDD) paradigm, the framework iteratively refines the synthesized code based on execution feedback from a symbolic executor, ensuring strict adherence to the underlying rules. Experimental results demonstrate that our method significantly reduces rule omission and logical errors compared to baselines such as directly prompting LLMs for results and code generation under full rule sets, providing a transparent and traceable reasoning process for complex real-world scenarios.

1 Introduction

1.1 研究背景

近年来，大语言模型（LLMs）在自然语言处理领域展现出了卓越的能力，但在面对现实世界中复杂的规则引导推理（Rule-Guided Reasoning）时，依然表现出明显的局限性。以 RULEARENA 数据集中的场景为例，无论是计算航空公司行李费、判断NBA球员交易是否符合劳资协议，还是处理严苛的税务审计计算，都需要模型

具备极高的逻辑严密性和计算精确度。在这些现实场景中，任何微小的规则理解偏差或计算错误都可能导致错误，因此，研究如何提升模型在复杂规则下的推理能力具有极重要的意义。

1.2 纯神经推理的局限性

为了推动这一领域的研究，本文深入探讨了包含上述三个领域的复杂规则数据集。然而，实验发现现有的最先进模型（SOTA）在这些任务上的表现均不尽如人意。通过分析，我们总结出其失败的核心原因在于：模型没有对必须复杂推理任务实现规则的精准理解与绝对遵守。具体表现为以下几点：

规则规模庞大：如税务相关规则有上千行，模型难以在长上下文中保持关注；

规则记忆与幻觉：根据我们的实验分析，随着规则数量增加，模型会对规则出现明显的记忆偏差，从而降低解答规则所对应的问题的准确率。

理解偏差：纯神经模型在解析自然语言规则时，经常产生逻辑偏差，往往需要繁琐的人工干预才能纠正。

1.3 神经符号方法

现有的纯生成式神经方法（Neural-only）本质上是基于概率的预测，很难在处理确定性极强的逻辑规则时保证零误差。因此，我们认为将符号系统（Symbolic）引入推理过程是必然选择。

规则天然地可以转化为确定性的编程代码或逻辑符号。在符号表达正确的前提下，程序执行可以完全避免神经模型在推理过程中的记忆损耗和随机性幻觉。基于此，我们提出了一种神经+符号（Neuro-Symbolic）多智能体框架。在该系统中，神经智能体负责将复杂的自然语言规则解析并转化为可执行的符号代码，而符号智能体负责执行逻辑并反馈结果。

这种神经符号协作模式的优势在于：它将 LLM 从沉重的逻辑计算中解放出来，让其回归到擅长的语义理

<p>解上，而将严密的逻辑推导交给稳定、透明且可解释的代码系统。</p> <h2>1.4 分步生成与 Case 驱动校验</h2> <p>然而，实现这种结合面临巨大的挑战：如何确保 LLM 生成的代码（符号）是完全正确的？为此，我们提出了基于逻辑分割的步骤化生成与自动边界 Case 校验机制。系统会要求模型自动根据规则生成一系列临界情况（Boundary Cases）作为验证集，并在代码空间中进行迭代搜索和调试。只有通过了全量校验的代码才会被视为最终的推理工具，从而从根本上保证了符号的正确性。</p> <p>实验结果表明，该机制显著提升了符号代码的生成质量。在与 GPT-4o 和 Qwen-2.5 72B 的对比实验中，我们的方法取得了性能提升。例如，在挑战性极大的 Airline 任务中，GPT-4o 在全量规则下的准确率几乎为 0，而本方法将其提升至了 0.83。</p> <p>本文的主要贡献总结如下：</p> <ul style="list-style-type: none"> 提出了一端到端的神经+符号多智能体推理框架，有效解决了 LLM 在复杂规则下的幻觉问题； 引入了分步生成与自动 Case 校验机制，利用模型自动生成的验证集在代码空间进行搜索，确保了符号逻辑的确定性； 在多个现实世界复杂任务上取得了显著的性能提升，验证了框架的鲁棒性和实用性。 	<p>自然语言翻译为符号公式（如一阶逻辑），并交由确定性求解器（如 Z3 或 Pyke）处理，从而保证了逻辑的严密性。ProofWriter [?] 则探索了模型生成证明链的能力，通过迭代生成一阶段蕴含（implication）来构建完整的逻辑证明。这类工作证明了符号逻辑系统在修正神经模型逻辑缺陷方面的巨大潜力。</p> <h2>2.3 智能体协作与工具增强</h2> <p>多智能体框架通过角色分工进一步提升了复杂问题的处理效率。DyLAN [Liu et al., 2024] 提出了动态智能体网络，根据任务贡献度自动筛选智能体团队进行协作。在工具利用方面，ChatCoT [Chen et al., 2023] 将思维链（CoT）推理建模为多轮对话，使模型能以更自然的方式调用外部工具（如 Python 解释器）。这些框架为复杂规则下的分工协作提供了范式参考。</p> <h2>2.4 本工作与现有工作的区别</h2> <h3>场景与数据集</h3> <p>现有工作如 ComplexBench 或 RuleEval 多侧重于人工构建的、规则条数较少的通用逻辑约束。相比之下，我们的 RULEARENA 数据集来源于真实的复杂领域（如 NBA 劳资协议、美国税务法规和民航行李费），其规则体量大（通常包含数百条相互关联的条款）、专业性强，对模型的长上下文管理和精准规则定位能力提出了更高挑战。</p> <h3>方法与框架</h3> <p>不同于 LOGIC-LM 等依赖单一转换的工作，我们提出了一个神经+符号的多智能体系统。该框架的核心创新在于引入了自动边界 Case 校验机制：针对复杂规则中难以通过人工覆盖的逻辑死角，系统能够自动生成边界测试案例作为“验证集”，对生成的符号代码进行在线回测和修正。这种闭环校验机制弥补了现有神经符号系统中“模型翻译代码后缺乏自审”的缺陷，显著提升了在严苛规则场景下的推理准确率。</p> <h2>3 Methodology</h2> <h3>3.1 任务建模与形式化定义</h3> <p>我们将复杂规则下的推理任务定义为三元组 (R,P,A)。</p> <p>R (Rules): 由数千字构成的自然语言规则集，包含显式定义的法律条文或行业规定，以及隐式存在的逻辑结构。</p>
--	--

<p>P (Problem): 包含具体数值、实体和约束条件的待解问题。</p> <p>A (Answer): 在 R 约束下, 针对 P 产生的唯一合法结果。</p> <p>传统的推理范式 $LLM(R, P) \rightarrow A$ 在 R 的长度增加时, 会因为注意力机制的稀释和上下文窗口的逻辑拥挤而产生幻觉。本方法引入符号层 S (Python 脚本), 将过程转化为:</p> <p>分段规则解构: 将 R 拆解为 $\{r_1, r_2, \dots, r_n\}$</p> <p>符号逻辑构建: 对于每个子任务, 通过 MAS 生成相应的符号函数 f_i。</p> <p>确定性执行: 最终结果 A 通过串联执行函数流 $\prod_{i=1}^n f_i(P)$ 获得。</p>	<h2>4 Experiments</h2> <h3>4.1 实验设置与数据集</h3> <p>我们在三个具有代表性的复杂规则领域开展了主实验:</p> <ul style="list-style-type: none"> Airline (航空行李费): 涉及复杂的航线区域映射与超重/超尺寸阶梯计费, 任务拆分为 9 步。 Tax (美国税务): 包含多表关联计算与累进税率逻辑, 任务拆分为 14 步。 NBA (劳资协议交易): 涉及硬帽限制、TPE 匹配等极端复杂的金融与合同约束, 任务拆分为 8 步。 <h3>4.2 规则规模对模型能力的影响</h3> <p>在开展主实验前, 我们探究了规则长度对于模型回答效果的简单实验。</p> <p>设计了 Airline 数据集上的简单问题, 在不断增加干扰规则的情况下考察模型对于相同问题的回答, 在不断增加规则长度的情况下要求模型回答规则对应的问题, 发现准确率均在一定程度的下降。</p> <p>对于 Airline 数据集的前一百个问题, 我们通过人工的方式找到了全量规则文本中与回答问题最相关的规则子集。发现在仅提供最相关的规则子集的情况下模型的回答会与真实答案更接近。</p> <p>这证明了将规则拆解的必要性。</p> <h3>4.3 主实验结果分析</h3> <p>我们将本方法与 0-shot 纯推理以及“全部规则不拆分”的基线进行了对比 (模型使用 Qwen-2.5 72B 与 GPT-4o)。实验数据显示:</p> <ul style="list-style-type: none"> 在 Tax 数据集上, 本方法在 GPT-4o 下达到了 100 在最复杂的 NBA 协议中, 本方法通过分段生成 Python 脚本, 解决了纯神经模型无法处理的 TPE 三级限额计算等硬逻辑瓶颈。 结论: 实验定性地证明, 通过将复杂规则“分而治之”并转化为可执行代码, 能有效消除模型在长程推理中的计算漂移。 <h3>4.4 案例分析: 从失败到成功的闭环</h3> <h3>4.5 效率损耗分析与工程瓶颈</h3> <p>我们对系统的推理损耗进行了定量化评估, 并得出以下定性结论:</p> <p>保险溢价成本: 虽然 MAS 方法平均产生了约 8.2 倍的调用损耗, 但这一成本换取了一定的逻辑确定性, 自动修复了大量因路径解析、API 参数或逻辑细节导致的潜在错误。</p>
---	--

工程复杂度瓶颈: 实验发现, 系统主要的效率损耗 230 并非源于复杂的算法计算 (如累进税率), 而是源于对 231 “物理世界”的摸索, 如处理跨文件引用或沙箱环境下的 232 权限限制。	271 22in 12-point bold font. On the following line(s) place the affil- iations.	272
协作演化趋势: 随着推理步骤的推进, 智能体表现 233 出角色合并的倾向。Code Agent 会自发承担部分测试 234 代码的编写工作, 使系统损耗趋于一个稳定的“地板水 235 平”(约 6 次调用), 展现了多智能体协作的自适应优化 236 能力。	273	
• If your track requires submissions to be anonymous, 238 they must be fully anonymized as discussed in the Mod- 239 ications for Blind Review subsection below; in this 240 case, Acknowledgements and Contribution Statement 241 sections are not allowed.	274	
• If your track requires non-anonymous submissions, you 243 should provide all author information at the time of 244 submission, just as for camera-ready papers (see be- 245 low); Acknowledgements and Contribution Statement 246 sections are allowed, but optional.	275	
• Submissions must include line numbers to facilitate 248 feedback in the review process. Enable line numbers by 249 uncommenting the command <code>\linenumbers</code> in the 250 preamble.	276	
• The limit on the number of content pages is <i>strict</i> . All 252 papers exceeding the limits will be desk rejected.	277	
Camera-Ready Papers	278	
The following instructions apply to camera-ready pa- pers:	279	
• Authors and affiliations are mandatory. Explicit self- 257 references are allowed. It is strictly forbidden to add 258 authors not declared at submission time.	280	
• Acknowledgements and Contribution Statement sec- 260 tions are allowed, but optional.	281	
• Line numbering must be disabled. To achieve this, com- 262 ment or disable <code>\linenumbers</code> in the preamble.	282	
• For some of the tracks, you can exceed the page limit by 264 purchasing extra pages.	283	
4.6 Title and Author Information	294	
Center the title on the entire width of the page in a 14- 267 point bold font. The title must be capitalized using Title 268 Case. For non-anonymous papers, author names and affili- 269 ations should appear below the title. Center author name(s)	295	
Author Names	296	
Each author name must be followed by: • A newline <code>\</code> command for the last author. • An <code>\And</code> command for the second to last author. • An <code>\and</code> command for the other authors.	297	
Affiliations	298	
After all authors, start the affiliations section by using the <code>\affiliations</code> command. Each affiliation must be terminated by a newline <code>\</code> command. Make sure that you include the newline after the last affiliation, too.	299	
Mapping Authors to Affiliations	300	
If some scenarios, the affiliation of each author is clear without any further indication (<i>e.g.</i> , all authors share the same affiliation, all authors have a single and different affiliation). In these situations you don't need to do anything special.	301	
In more complex scenarios you will have to clearly in- dicate the affiliation(s) for each author. This is done by using numeric math superscripts $\{^i, j, \dots\}$. You must use num- bers, not symbols, because those are reserved for footnotes in this section (should you need them). Check the authors defi- nition in this example for reference.	302	
Emails	303	
This section is optional, and can be omitted entirely if you prefer. If you want to include e-mails, you should either include all authors' e-mails or just the contact author(s)' ones.	304	
Start the e-mails section with the <code>\emails</code> command. After that, write all emails you want to include separated by a comma and a space, following the order used for the authors (<i>i.e.</i> , the first e-mail should correspond to the first author, the second e-mail to the second author and so on).	305	
You may “contract” consecutive e-mails on the same do- main as shown in this example (write the users' part within curly brackets, followed by the domain name). Only e- mails of the exact same domain may be contracted. For instance, you cannot contract “ <code>person@example.com</code> ” and “ <code>other@test.example.com</code> ” because the domains are differ- ent.	306	
Modifications for Blind Review	307	
When submitting to a track that requires anonymous submissions, in order to make blind reviewing possible, au- thors must omit their names, affiliations and e-mails. In place	308	

of names, affiliations and e-mails, you can optionally provide³⁵⁵

the submission number and/or a list of content areas. When³⁵⁶ referring to one's own work, use the third person rather than the first person. For example, say, "Previously, Gottlob [?] has shown that...", rather than, "In our previous work [?], we have shown that..." Try to avoid including any information in the body of the paper or references that would identify the authors or their institutions, such as acknowledgements. Such information can be added post-acceptance to be included in the camera-ready version. Please also make sure that your paper metadata does not reveal the authors' identities.

4.7 Abstract

Place the abstract at the beginning of the first column 3" from the top of the page, unless that does not leave enough room for the title and author information. Use a slightly smaller width than in the body of the paper. Head the abstract with "Abstract" centered above the body of the abstract in a 12-point bold font. The body of the abstract should be in the same font as the body of the paper.

The abstract should be a concise, one-paragraph summary describing the general thesis and conclusion of your paper. A reader should be able to learn the purpose of the paper and the reason for its importance from the abstract. The abstract should be no more than 200 words long.

4.8 Text

The main body of the text immediately follows the abstract. Use 10-point type in a clear, readable font with 1-point leading (10 on 11).

Indent when starting a new paragraph, except after major headings.

4.9 Headings and Sections

When necessary, headings should be used to separate major sections of your paper. (These instructions use many headings to demonstrate their appearance; your paper should have fewer headings.). All headings should be capitalized using Title Case.

Section Headings

Print section headings in 12-point bold type in the style shown in these instructions. Leave a blank space of approximately 10 points above and 4 points below section headings. Number sections with Arabic numerals.

Subsection Headings

Print subsection headings in 11-point bold type. Leave a blank space of approximately 8 points above and 3 points below subsection headings. Number subsections with the section number and the subsection number (in Arabic numerals) separated by a period.

Subsubsection Headings

Print subsubsection headings in 10-point bold type. Leave a blank space of approximately 6 points above subsubsection headings. Do not number subsubsections.

Titled paragraphs. You should use titled paragraphs if and only if the title covers exactly one paragraph. Such paragraphs should be separated from the preceding content by at least 3pt, and no more than 6pt. The title should be in 10pt bold font and to end with a period. After that, a 1em horizontal space should follow the title before the paragraph's text.

In L^AT_EX titled paragraphs should be typeset using

```
\paragraph{Title.} text .
```

4.10 Special Sections

Appendices

You may move some of the contents of the paper into one or more appendices that appear after the main content, but before references. These appendices count towards the page limit and are distinct from the supplementary material that can be submitted separately through CMT. Such appendices are useful if you would like to include highly technical material (such as a lengthy calculation) that will disrupt the flow of the paper. They can be included both in papers submitted for review and in camera-ready versions; in the latter case, they will be included in the proceedings (whereas the supplementary materials will not be included in the proceedings). Appendices are optional. Appendices must appear after the main content. Appendix sections must use letters instead of Arabic numerals. In L^AT_EX, you can use the \appendix command to achieve this followed by \section{Appendix} for your appendix sections.

Ethical Statement

Ethical Statement is optional. You may include an Ethical Statement to discuss the ethical aspects and implications of your research. The section should be titled *Ethical Statement* and be typeset like any regular section but without being numbered. This section may be placed on the References pages.

Use

```
\section*{Ethical Statement}
```

Acknowledgements

Acknowledgements are optional. In the camera-ready version you may include an unnumbered acknowledgments section, including acknowledgments of help from colleagues, financial support, and permission to publish. This is not allowed in the anonymous submission. If present, acknowledgements must be in a dedicated, unnumbered section appearing after all regular sections but before references. This section may be placed on the References pages.

Use

```
\section*{Acknowledgements}
```

to typeset the acknowledgements section in L^AT_EX.

Contribution Statement

Contribution Statement is optional. In the camera-ready version you may include an unnumbered Contribution Statement section, explicitly describing the contribution of each of the co-authors to the paper. This is not allowed in the anonymous submission. If present, Contribution Statement must be in a dedicated, unnumbered section appearing after all regular sections but before references. This section may be placed on the References pages.

Use

```
\section*{Contribution Statement}
```

to typeset the Contribution Statement section in L^AT_EX.

References

The references section is headed “References”, printed in the same style as a section heading but without a number. A sample list of references is given at the end of these instructions. Use a consistent format for references. The reference list should not include publicly unavailable work.

Order of Sections

Sections should be arranged in the following order:

1. Main content sections (numbered)
2. Appendices (optional, numbered using capital letters)
3. Ethical statement (optional, unnumbered)
4. Acknowledgements (optional, unnumbered)
5. Contribution statement (optional, unnumbered)
6. References (required, unnumbered)

4.11 Citations

Citations within the text should include the author’s last name and the year of publication, for example [?]. Append lowercase letters to the year in cases of ambiguity. Treat multiple authors as in the following examples: [?] or [?] (for more than two authors) and [?] (for two authors). If the author portion of a citation is obvious, omit it, e.g., Nebel [?]. Collapse multiple citations as follows: [?; ?].

4.12 Footnotes

Place footnotes at the bottom of the page in a 9-point font. Refer to them with superscript numbers.¹ Separate them from the text by a short line.² Avoid footnotes as much as possible; they interrupt the flow of the text.

5 Illustrations

Place all illustrations (figures, drawings, tables, and photographs) throughout the paper at the places where they are first discussed, rather than at the end of the paper.

They should be floated to the top (preferred) or bottom of the page, unless they are an integral part of your narrative flow. When placed at the bottom or top of a page, illustrations may run across both columns, but not when they appear inline.

Illustrations must be rendered electronically or scanned and placed directly in your document. They should be cropped outside L^AT_EX, otherwise portions of the image could reappear during the post-processing of your paper. When possible, generate your illustrations in a vector format. When using bitmaps, please use 300dpi resolution at least. All illustrations should be understandable when printed in black and white, albeit you can use colors to enhance them. Line weights should be 1/2-point or thicker. Avoid screens and superimposing type on patterns, as these effects may not reproduce well.

Number illustrations sequentially. Use references of the following form: Figure 1, Table 2, etc. Place illustration numbers and captions under illustrations. Leave a margin of 1/4-inch around the area covered by the illustration and caption. Use 9-point type for captions, labels, and other text in illustrations. Captions should always appear below the illustration.

¹This is how your footnotes should appear.

²Note the line separating these footnotes from the text.

Scenario	δ	Runtime
Paris	0.1s	13.65ms
Paris	0.2s	0.01ms
New York	0.1s	92.50ms
Singapore	0.1s	33.33ms
Singapore	0.2s	23.01ms

表 1: Latex default table

Scenario	δ (s)	Runtime (ms)
Paris	0.1	13.65
	0.2	0.01
New York	0.1	92.50
Singapore	0.1	33.33
	0.2	23.01

表 2: Booktabs table

6 Tables

Tables are treated as illustrations containing data. Therefore, they should also appear floated to the top (preferably) or bottom of the page, and with the captions below them.

If you are using L^AT_EX, you should use the `booktabs` package, because it produces tables that are better than the standard ones. Compare Tables 1 and 2. The latter is clearly more readable for three reasons:

1. The styling is better thanks to using the `booktabs` rulers instead of the default ones.
2. Numeric columns are right-aligned, making it easier to compare the numbers. Make sure to also right-align the corresponding headers, and to use the same precision for all numbers.
3. We avoid unnecessary repetition, both between lines (no need to repeat the scenario name in this case) as well as in the content (units can be shown in the column header).

7 Formulas

IJCAI's two-column format makes it difficult to typeset long formulas. A usual temptation is to reduce the size of the formula by using the `small` or `tiny` sizes. This doesn't work correctly with the current L^AT_EX versions, breaking the line spacing of the preceding paragraphs and title, as well as the equation number sizes. The following equation demonstrates the effects (notice that this entire paragraph looks badly formatted, and the line numbers no longer match

the text):

$$x = \prod_{i=1}^n \sum_{j=1}^n j_i + \prod_{i=1}^n \sum_{j=1}^n i_j + \prod_{i=1}^n \sum_{j=1}^n j_i + \prod_{i=1}^n \sum_{j=1}^n i_j + \prod_{i=1}^n \sum_{j=1}^n j_i \quad (1)$$

Reducing formula sizes this way is strictly forbidden. We strongly recommend authors to split formulas in multiple lines when they don't fit in a single line. This is the easiest approach to typeset those formulas and provides the most readable output

$$\begin{aligned} x = & \prod_{i=1}^n \sum_{j=1}^n j_i + \prod_{i=1}^n \sum_{j=1}^n i_j + \prod_{i=1}^n \sum_{j=1}^n j_i + \prod_{i=1}^n \sum_{j=1}^n i_j + \\ & + \prod_{i=1}^n \sum_{j=1}^n j_i. \end{aligned} \quad (2)$$

If a line is just slightly longer than the column width, you may use the `resizebox` environment on that equation. The result looks better and doesn't interfere with the paragraph's line spacing:

$$x = \prod_{i=1}^n \sum_{j=1}^n j_i + \prod_{i=1}^n \sum_{j=1}^n i_j + \prod_{i=1}^n \sum_{j=1}^n j_i + \prod_{i=1}^n \sum_{j=1}^n i_j + \prod_{i=1}^n \sum_{j=1}^n j_i. \quad (3)$$

This last solution may have to be adapted if you use different equation environments, but it can generally be made to work. Please notice that in any case:

- Equation numbers must be in the same font and size as the main text (10pt).
- Your formula's main symbols should not be smaller than small text (9pt).

For instance, the formula

$$x = \prod_{i=1}^n \sum_{j=1}^n j_i + \prod_{i=1}^n \sum_{j=1}^n i_j + \prod_{i=1}^n \sum_{j=1}^n j_i + \prod_{i=1}^n \sum_{j=1}^n i_j + \prod_{i=1}^n \sum_{j=1}^n j_i + \prod_{i=1}^n \sum_{j=1}^n i_j \quad (4)$$

would not be acceptable because the text is too small.

8 Examples, Definitions, Theorems and Similar

Examples, definitions, theorems, corollaries and similar must be written in their own paragraph. The paragraph must be separated by at least 2pt and no more than 5pt from the preceding and succeeding paragraphs. They must begin with the kind of item written in 10pt bold font followed by their number (e.g.: **Theorem 1**), optionally followed by a title/summary between parentheses in non-bold font and ended with a period (in bold). After that the main body of the item follows, written in 10 pt italics font (see below for examples).

In \LaTeX we strongly recommend that you define environments for your examples, definitions, propositions, lemmas, corollaries and similar. This can be done in your \LaTeX preamble using `\newtheorem` – see the source of this document for examples. Numbering for these items must be global, not per-section (e.g.: Theorem 1 instead of Theorem 6.1).

Example 1 (How to write an example). *Examples should be written using the example environment defined in this template.*

Theorem 1. *This is an example of an untitled theorem.*

You may also include a title or description using these environments as shown in the following theorem.

Theorem 2 (A titled theorem). *This is an example of a titled theorem.*

9 Proofs

Proofs must be written in their own paragraph(s) separated by at least 2pt and no more than 5pt from the preceding and succeeding paragraphs. Proof paragraphs should start with the keyword “Proof.” in 10pt italics font. After that the proof follows in regular 10pt font. At the end of the proof, an unfilled square symbol (qed) marks the end of the proof.

In \LaTeX proofs should be typeset using the `\proof` environment.

证明. This paragraph is an example of how a proof looks like using the `\proof` environment. \square

10 Algorithms and Listings

Algorithms and listings are a special kind of figures. Like all illustrations, they should appear floated to the top (preferably) or bottom of the page. However, their caption should appear in the header, left-justified and enclosed between horizontal lines, as shown in Algorithm 1. The algorithm body should be terminated with another horizontal line. It is up to the authors to decide whether to show line numbers or not, how to format comments, etc.

In \LaTeX algorithms may be typeset using the `algorithm` and `algorithmic` packages, but you can also use one of the many other packages for the task.

535 **Algorithm 1** Example algorithm

536 **Input:** Your algorithm’s input

537 **Parameter:** Optional list of parameters

538 **Output:** Your algorithm’s output

539 1: Let $t = 0$.

540 2: **while** condition **do**

3: Do some action.

4: **if** conditional **then**

5: Perform task A.

6: **else**

7: Perform task B.

8: **end if**

9: **end while**

10: **return** solution

11 \LaTeX and Word Style Files

The \LaTeX and Word style files are available on the IJCAI-ECAI 26 website, <https://2026.ijcai.org/>. These style files implement the formatting instructions in this document.

The \LaTeX files are `ijcai26.sty` and `ijcai26.tex`, and the Bib \TeX files are named `.bst` and `ijcai26.bib`. The \LaTeX style file is for version 2e of \LaTeX , and the Bib \TeX style file is for version 0.99c of Bib \TeX (*not* version 0.98i).

The Microsoft Word style file consists of a single file, `ijcai26.docx`.

These Microsoft Word and \LaTeX files contain the source of the present document and may serve as a formatting sample.

Further information on using these styles for the preparation of papers for IJCAI-ECAI 26 can be obtained by contacting `proceedings@ijcai.org`.

Ethical Statement

There are no ethical issues.

Acknowledgments

The preparation of these instructions and the \LaTeX and Bib \TeX files that implement them was supported by Schlumberger Palo Alto Research, AT&T Bell Laboratories, and Morgan Kaufmann Publishers. Preparation of the Microsoft Word file was supported by IJCAI. An early version of this document was created by Shirley Jowell and Peter F. Patel-Schneider. It was subsequently modified by Jennifer Ballentine, Thomas Dean, Bernhard Nebel, Daniel Pagenstecher,

Kurt Steinkraus, Toby Walsh, Carles Sierra, Marc Pujol-Gonzalez, Francisco Cruz-Mencia and Edith Elkind. 601
602

参考文献 603

[Chen *et al.*, 2023] Zhipeng Chen, Kun Zhou, Beichen Zhang, Zheng Gong, Wayne Xin Zhao, and Ji-Rong Wen. 604
605

Chatcot: Tool-augmented chain-of-thought reasoning on 606
chat-based large language models. In *arXiv preprint arXiv:2305.14323*, 2023. 607
608

[Liu *et al.*, 2024] Zijun Liu, Yanzhe Zhang, Peng Li, Yang Liu, and Diyi Yang. Dynamic llm-powered agent network 609
for task-oriented agent collaboration. In *Proceedings of 610*
the 1st Conference on Language Modeling (COLM), 2024. 611
612

[Pan *et al.*, 2023] Liangming Pan, Alon Albalak, Xinyi Wang, and William Yang Wang. Logic-lm: Empowering 613
large language models with symbolic solvers for faithful 614
logical reasoning. In *Proceedings of the 2023 Conference 615*
on Empirical Methods in Natural Language Processing 616
(EMNLP), 2023. 617
618

[Sun *et al.*, 2024] Wangtao Sun, Chenxiang Zhang, Xueyou Zhang, Xuanqing Yu, Ziyang Huang, Haotian Xu, Pei Chen, Shizhu He, Jun Zhao, and Kang Liu. Beyond 619
instruction following: Evaluating inferential rule fol- 620
lowing of large language models. In *arXiv preprint arXiv:2407.08440*, 2024. 621
622
623

[Wang *et al.*, 2023] Bailin Wang, Zhaofeng Wu, Linlu Qiu, Alexis Ross, Ekin Akyürek, Najoung Kim, Jacob Andreas, 625
and Yoon Kim. Reasoning or reciting? exploring the 626
capabilities and limitations of language models through 627
counterfactual tasks. In *arXiv preprint arXiv:2307.02477*, 628
2023. 629
630

[Wen *et al.*, 2024] Bosi Wen, Pei Ke, Xiaotao Gu, Lindong Wu, Hao Huang, Jinfeng Zhou, Wenchuang Li, Binxin Hu, Wendy Gao, Jiaxin Xu, et al. Benchmarking complex instruction-following with multiple constraints composition. In *Proceedings of the 38th Conference on Neural 631*
Information Processing Systems (NeurIPS), 2024. 632
633
634
635
636

[Xu *et al.*, 2024] Can Xu, Qingfeng Sun, Kai Zheng, Xiubo Geng, Pu Zhao, Jiazhan Feng, Chongyang Tao, Qingwei Lin, and Dixin Jiang. Wizardlm: Empowering large pre-trained language models to follow complex instructions. In *International Conference on Learning Representations 637*
(ICLR), 2024. 638
639
640
641
642