

思维对话：生成特定领域信息的协作多代理系统

Christodoulos Constantinides¹, Shuxin Lin², Nianjun Zhou², Dhaval Patel²

¹IBM

²IBM Research

{ christodoulos.constantinides@, shuxin.lin@, jzhou@us., pateldha@us. } ibm.com,

Abstract

本文介绍了一种名为 Chat-of-Thought 的新型多智能体系统，旨在促进工业资产的失效模式及影响分析（FMEA）文件的生成。Chat-of-Thought 利用了多个基于大型语言模型（LLM）的智能体，每个智能体扮演特定的角色，通过高级人工智能技术和动态任务路由来优化 FMEA 表格的生成和验证。该系统的一个重要创新是引入了“思维聊天”，通过动态、多角色驱动的讨论，实现内容的迭代改进。本研究探讨了工业设备监控的应用领域，突出显示了主要挑战，并展示了 Chat-of-Thought 通过互动的、模板驱动的工作流程和环境感知的智能体协作来应对这些挑战的潜力。

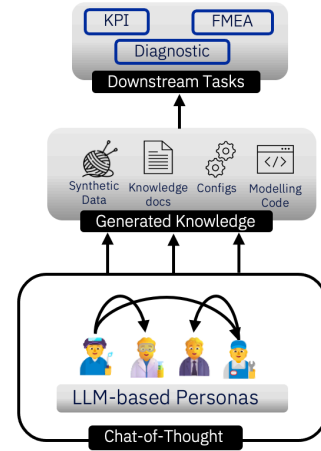


Figure 1: 聊天思维概述。

1 引言

在现代数据驱动的行业，团队经常依赖多样化的专业知识来提取可操作的洞察并优化运营。主题专家、数据科学家和工程师之间的协作努力在分析复杂系统、识别风险和改善决策方面起着关键作用。然而，这些过程资源密集并且需要大量协调。为了解决这些挑战，我们提出了一个新颖的框架，Chat-of-Thought，它利用大型语言模型（LLMs）[Achiam et al., 2023]、[Touvron et al., 2023]、[Jiang et al., 2023] 的角色扮演能力来模拟协作、多代理环境 [Park et al., 2023]，以生成特定领域的知识。

Chat-of-Thought 系统可以创建虚拟人物，每个人物代表一个特定的专业领域。这些人物协同工作，以生成假设的文件和见解，而无需人工干预。通过整合来自多种信息来源的输入——如失效模式与影响分析（FMEA）、关键绩效指标（KPIs）、诊断数据、YAML 配置、合成数据集和建模代码——该框架促进了以自动化和可扩展的方式探索复杂场景。

该框架的一个关键应用是 FMEA 的自动化，FMEA 是一种系统的可靠性工程方法，用于识别和减轻工业系统中的潜在失效模式。手动创建 FMEA 文档劳动强度大且需要多学科专业知识，使其成为自动化的理想选择。我们的系统 Chat-of-Thought 采用具有专门角色的多代理架构来简化 FMEA 的生成，确保准确性和可扩展性，同时显著减少所需的工作量。

通过将模拟角色的协作潜力与多样化的数据输入相结合，Chat-of-Thought 系统展示了一种变革性的流程优化、风险识别和决策制定方法。这项工作旨在展示大型

语言模型（LLMs）在变革传统工作流程方面的潜力，使行业能够利用人工智能提升可靠性和运营效率。

1.1 应用领域和问题场景

该系统专为管理复杂高价值资产的行业设计，包括泵、锅炉、冷水机组和类似设备。这些资产固有地容易因机械、电气和操作问题而发生故障，通常导致较长的停机时间和高昂的成本。失效模式和影响分析（FMEA）文档的手动创建是可靠性工程中的一项关键任务，但因缺乏可扩展性、一致性和验证延迟等限制而受到影响，尤其是在涉及新型或不常见资产的超出范围（OOS）情景中。该领域的主要挑战包括：

- 对高度专业化领域特定专业知识的要求。
- 准确地理解多样化资产类别及其相关的失效机制的上下文。
- 简化和高效的 FMEA 数据验证，以最小化主题专家（SMEs）的工作量。

2 系统架构和关键组件

所提出的 Chat-of-Thought 系统采用了由大型语言模型（LLMs）驱动的协作多智能体框架。系统中的每个智能体都被设计了预定义的角色、技能和上下文系统消息，使其能够模拟特定领域的专业知识。该架构被分为以下关键阶段：

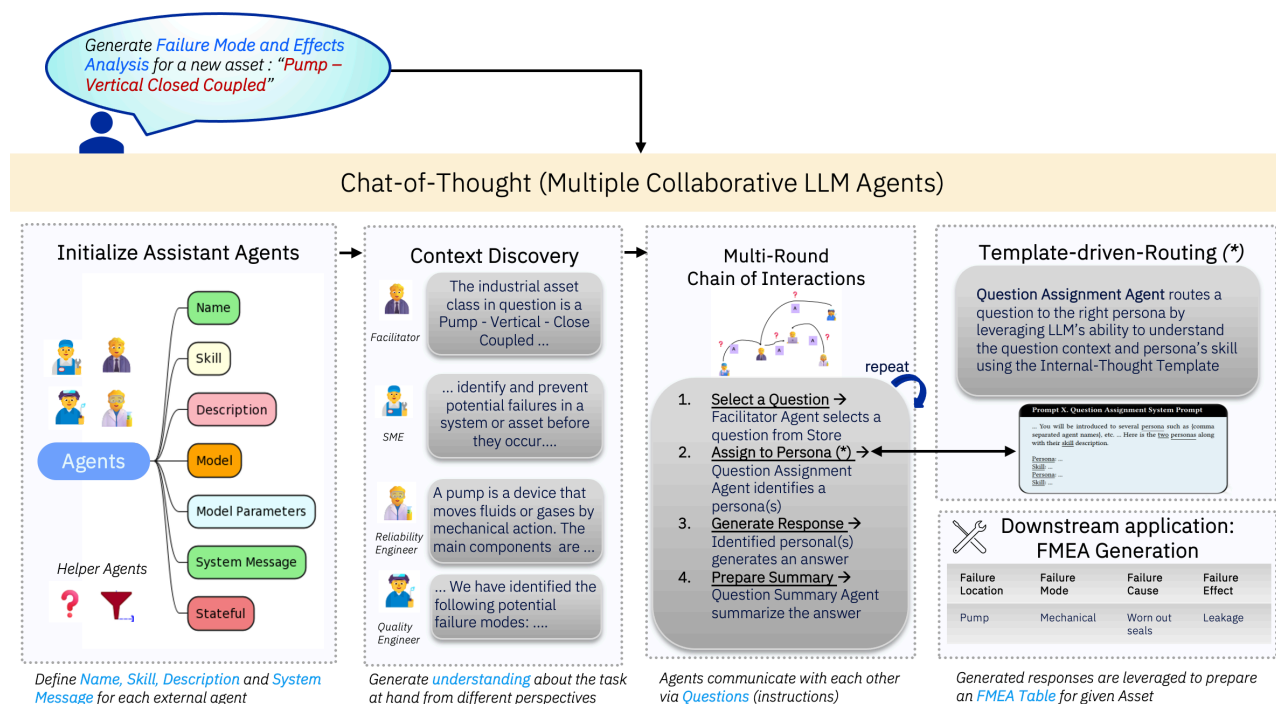


Figure 2: 聊天思维系统架构概述

2.1 助手代理的初始化

个代理被初始化，带有预定义的属性，以模拟 FMEA 生成中所需的特定角色。这些角色包括但不限于：每个代理 A_i 被表示为一个元组 (R_i, S_i, M_i) ，其中 R_i 表示角色， S_i 表示技能集， M_i 指定引导代理行为的系统消息。

2.2 上下文发现

该系统首先识别资产类别（例如，泵 - 垂直紧耦合）及其操作参数。相关的失效模式、原因及关联的影响会动态地从特定领域的知识库和历史数据中提取。该阶段确保分析针对被检查资产的特定上下文量身定制。

2.3 多轮相互作用链

Chat-of-Thought 的核心创新在于其 Chat-of-Thought 机制，其中代理通过迭代的、多角色驱动的讨论协作改进输出。这种方法通过促进跨越专业角色的动态、上下文感知的对话，超越了传统的 Chain-of-Thought 框架。该过程包括以下步骤：

1. 问题选择：促进者代理从预定义的问题库 Q_B 中选择一个问题 Q 。
2. 动态角色分配：问题分配代理根据上下文理解 C 和技能匹配 S_i 将选定的问题路由到最合适的角色 A_i 。
3. 回应生成：指定的角色生成回应 R ，结合其领域特定的专业知识和 LLM 能力。
4. 总结汇总：问题总结代理将响应汇聚成一个连贯且可操作的总结，以供后续处理。

为了确保高质量的输出，系统采用了一种基于模板的路由机制。每个模板 T 定义了将问题路由到合适角色的

结构化指南。这些模板会动态调整，以保持上下文的相关性并优化所生成 FMEA 数据的质量。通过利用这种模板驱动的方法，Chat-of-Thought 在适应各种场景的细微差别的同时，确保了交互的一致性。

这种架构通过强调协作智能和角色特定的专业知识，使得各行业能够在保持精确性和可扩展性的同时，有效地自动化劳动密集型的 FMEA 任务。

2.4 质量检查

在每轮结束时，有一个质量控制机制，使用预训练的分类器过滤掉无用的问题，使用自 BLEU 得分 [Papineni et al., 2002] 阈值去除重复的问题/答案。

3 人工智能技术与创新

Chat-of-Thought 集成了先进的 AI 技术，以简化 FMEA 生成并提升复杂任务的性能。其核心在于利用大型语言模型 (LLMs) 的强大能力进行领域特定推理、自然语言理解和生成高质量输出。任务分配通过上下文感知路由机制来引导，其中内部思维模板将代理技能与查询的具体要求对齐。多代理协作确保输出的迭代改进，而有状态的设计则使代理可以高效地管理多步骤和依赖上下文的过程。

该系统通过一个多阶段过程运行，该过程结合了失效模式分析、迭代学习和问答框架。初始阶段（第 1 轮）评估零样本学习能力，测试系统在缺乏任务特定示例时的基线性能。随后是上下文中学习（第 2 轮），其中源自失效模式及影响分析 (FMEA) 的场景指导响应生成，并结合人类反馈进行优化。

后续轮次建立在该基础上（第 3 轮）。链式互动技术在问题库和答案库之间引入反馈循环，从而实现迭代学习和持续改进。随机少样本学习方法（第 4 轮）增强了系

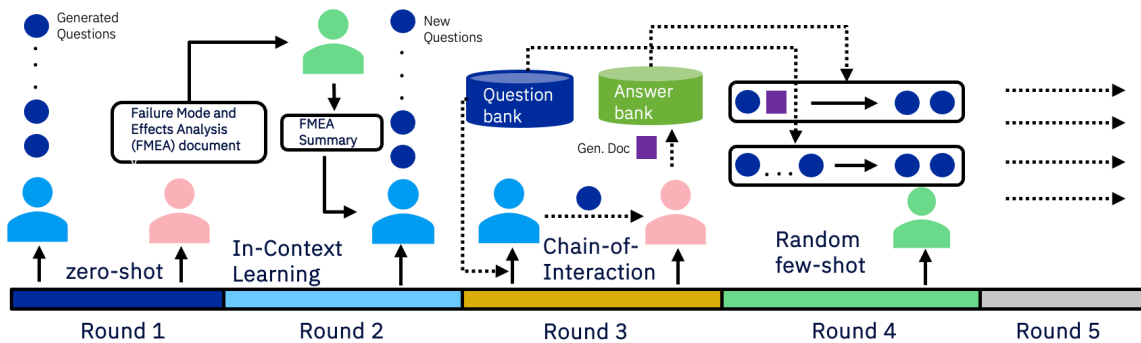


Figure 3: 回合推进。Chat-of-Thought 是一个灵活的框架，其中代理使用各种方法进行协作。

统从有限样本中普遍化知识的能力，使其能够适应新的或不常见的情境。这些迭代阶段最终形成一个强大且优化的系统，能够应对复杂的、特定领域的挑战。

4 结果和演示

思维对话的展示突显了其在各种操作阶段的能力。代理被初始化为不同的角色和技能，以协作生成 FMEA 内容。展示了思维对话过程，在这一过程中，多代理迭代讨论优化了输出并确保了上下文相关的响应。该系统还包括一个供专家审查的界面，从而实现对生成的 FMEA 表格的高效验证。

Chat-of-Thought 展示了其为标准和范围外资产生成详细且准确的 FMEA 表格的能力。专家的验证确认该系统能够可靠地识别失效模式、根本原因和潜在影响。这些输出有助于实施预防性维护策略，并在工业应用中提升整体可靠性规划。

在补充材料中，我们提供了一个视频来演示该系统。

References

- [Achiam et al., 2023] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. arXiv preprint arXiv:2303.08774, 2023.
- [Bhat et al., 2022] Shravya Bhat, Huy Nguyen, Steven Moore, John Stamper, Majd Sakr, and Eric Nyberg. Towards automated generation and evaluation of questions in educational domains. In Antonija Mitrovic and Nigel Bosch, editors, Proceedings of the 15th International Conference on Educational Data Mining, pages 701–704, Durham, United Kingdom, July 2022. International Educational Data Mining Society.
- [Cao and Wang, 2021] Shuyang Cao and Lu Wang. Controllable open-ended question generation with a new question type ontology, 2021.
- [Constantinides et al., 2025] Christodoulos Constantinides, Vivek Sharma, Shuxin Lin, Nianjun Zhou, Bharathi Chaudhury, and Dhaval Patel. Auto-q : Automated domain questions generation for industrial assets. In Towards Knowledgeable Foundation Models @ AAAI 2025 Workshop. Association for the Advancement of Artificial Intelligence, 2025.
- [Drori et al., 2023] Iddo Drori, Sarah J. Zhang, and etc Shuttleworth. From human days to machine seconds: Automatically answering and generating machine learning final exams. In SIGKDD, KDD '23, page 3947–3955, New York, NY, USA, 2023.
- [Gong et al., 2022] Huanli Gong, Liangming Pan, and Hengchang Hu. KHANQ: A dataset for generating deep questions in education. In Proceedings of the 29th International Conference on Computational Linguistics, pages 5925–5938, Gyeongju, Republic of Korea, October 2022. International Committee on Computational Linguistics.
- [Guanliang Chen and Houben, 2018] Claudia Hauff Guanliang Chen, Jie Yang and Geert-Jan Houben.

- Learningq: A large-scale dataset for educational question generation, 2018.
- [He et al., 2023] Zhankui He, Zhouhang Xie, and etc Jha. Large language models as zero-shot conversational recommenders. In CIKM, CIKM ' 23. ACM, October 2023.
- [IBM, 2024] IBM. Ibm generative ai python sdk (tech preview). <https://github.com/IBM/ibm-generative-ai>, 2024. Accessed 2024-01-02.
- [Jiang et al., 2023] Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. Mistral 7b. arXiv preprint arXiv:2310.06825, 2023.
- [Kitaev et al., 2019] Nikita Kitaev, Steven Cao, and Dan Klein. Multilingual constituency parsing with self-attention and pre-training, 2019.
- [Kojima et al., 2022] Takeshi Kojima, Shixiang (Shane) Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, Advances in Neural Information Processing Systems, volume 35, pages 22199–22213. Curran Associates, Inc., 2022.
- [Li et al., 2015] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. A diversity-promoting objective function for neural conversation models. arXiv preprint arXiv:1510.03055, 2015.
- [Li, 2023] Yucheng Li. An open source data contamination report for large language models, 2023.
- [Papineni et al., 2002] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics, pages 311–318, 2002.
- [Park et al., 2023] Joon Sung Park, Joseph O’Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. In Proceedings of the 36th annual acm symposium on user interface software and technology, pages 1–22, 2023.
- [Sanseviero et al., 2023] Omar Sanseviero, Lewis Tunstall, Philipp Schmid, Sourab Mangrulkar, Younes Belkada, and Pedro Cuenca. Mixture of experts explained, 2023.
- [Sun et al., 2023] Zhiqing Sun, Yikang Shen, Qinzhong Zhou, Hongxin Zhang, Zhenfang Chen, David Cox, Yiming Yang, and Chuang Gan. Principle-driven self-alignment of language models from scratch with minimal human supervision. In Thirty-seventh Conference on Neural Information Processing Systems, 2023.
- [Touvron et al., 2023] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971, 2023.
- [Uptake, 2024] Uptake. Asset Strategy Library, 2024. Accessed on 1 18, 2024.
- [Wang et al., 2022] Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A. Smith, Daniel Khashabi, and Hannaneh Hajishirzi. Self-instruct: Aligning language model with self generated instructions, 2022.
- [Wikipedia, 2023] Wikipedia. Cross-industry standard process for data mining — Wikipedia, the free encyclopedia. <http://en.wikipedia.org/w/index.php?title=Cross-industry%20standard%20process%20for%20data%20mining&oldid=1181901999>, 2023. [Online; accessed 18-December-2023].
- [Wu et al., 2023] Qingyun Wu, Gagan Bansal, and Jieyu Zhang etc. Autogen: Enabling next-gen llm applications via multi-agent conversation, 2023.