

【神经符号 AI 讲座】神经符号人工智能—2020 年以来的基础 研究进展与科学问题

编者按：本文编译自马里兰大学 Brandon Colelough 和 William Regli 的综述论文《Neuro-Symbolic AI in 2024: A Systematic Review》（根据 **Creative Commons License Attribution 4.0 International (CC BY 4.0)** 许可使用）。有兴趣的读者请参阅原文 **Colelough B C, Regli W. Neuro-Symbolic AI in 2024: A Systematic Review[J]. arXiv preprint arXiv:2501.05435, 2025.**

自 2020 年以来，神经符号人工智能领域的研究活动显著增长，这反映出人们对将符号与亚符号方法相结合以增强人工智能推理能力重要性的日益重视。神经符号人工智能的五大基础研究领域包括：**知识表示、学习与推断、逻辑与推理、可解释性与可信度、元认知**。近年来，神经符号人工智能在学习与推断方面取得了显著进展，然而，在可解释性与可信度以及元认知方面，仍存在显著的研究空白。通过跨学科研究来填补这些空白，对于推动神经符号人工智能领域向更智能、更可靠、更具情境感知能力的 AI 系统发展至关重要。

一、符号 AI、亚符号 AI 和神经符号 AI

1.1. 符号人工智能

符号人工智能（Symbolic AI）是人工智能的一个分支领域，与学习周围世界的内部符号表征有关，是“将某种形式的隐含人类知识转化为更形式化的和基于规则和逻辑的声明形式”¹。

1.2. 亚符号人工智能

相比之下，亚符号人工智能（Sub-Symbolic AI）不需要规则或符号表示作为输入，而是自己学习隐式数据表示的系统。亚符号人工智能包括机器学习、深度学习和生成式人工智能等方法，这些方法依赖于算法从原始数据中自动提取模式，以辨别关系并根据学习的表示进行预测。一些最早使用亚符号表示的人工智能系统的例子：多层感知机、Hopfield 网络、反向传播算法，以及一些最新的亚符号

¹ A. Dingli, Neuro-symbolic ai, 2023. Includes bibliographical references and index.

系统例子，如 GPT 模型、基于卷积神经网络的 YOLO 家族、DALLE 扩散模型等。

1.3.神经符号人工智能

近年兴起神经符号人工智能（Neuro-symbolic AI），试图结合符号 AI 和亚符号 AI 的优点，用逻辑规则约束深度学习模型，从数据中提取符号化知识，是“一种复合人工智能框架，通过融合符号人工智能和亚符号人工智能，创建具有推理能力的高级混合人工智能模型”²。神经符号人工智能的思想借鉴了丹尼尔·卡尼曼（Daniel Kahneman）《思考，快与慢》关于系统 1 和系统 2 思维的概念。其中，系统 1 是快速、直觉和并行的，类似于深度学习的能力；而系统 2 则是缓慢、谨慎和顺序性的，类似于符号推理。

二、神经符号 AI 的基础研究进展

从 1970 年到现在，仅 Google Scholar 就列出了 957 份出版物。图 1 显示了神经符号 AI 的研究从 2020 年开始呈指数级增长，到 2023 年达到顶峰（236 篇出版物）。

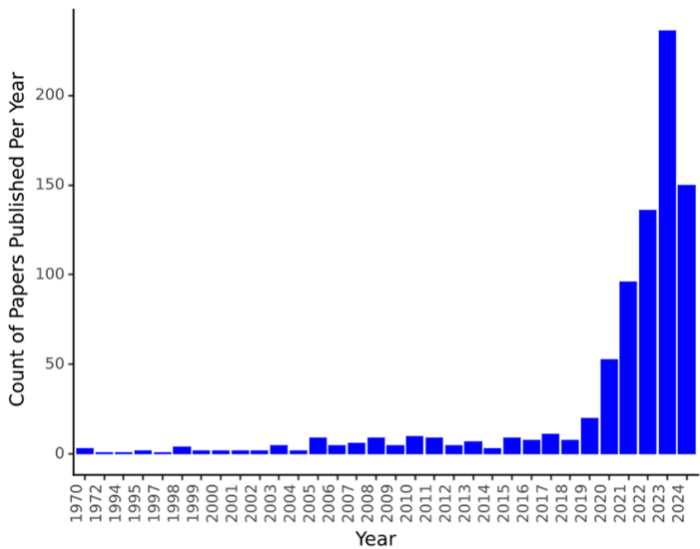


Figure 1: Histogram of publications per year for Neuro-Symbolic AI. The data was obtained through Google Scholar scraping, reflecting significant growth from 2020

图 1 神经符号人工智能每年发表的文章³(数据通过 Google Scholar 抓取获得的，反映了 2020 年以来的显著增长)

² A. d. Garcez, L. C. Lamb, Neurosymbolic ai: the 3rd wave, Artificial Intelligence Review 56 (2023) 12387–12406. doi:10. 1007/s10462-023-10448-w.

³ Colelough B C, Regli W. Neuro-Symbolic AI in 2024: A Systematic Review[J]. arXiv preprint arXiv:2501.05435, 2025.

神经符号人工智能的五大基础研究领域如下：

知识表示 (Knowledge Representation)：融合符号与神经表示，开发通用常识和特定领域的知识图谱，以实现更丰富的语义表达和知识组织。

学习与推断 (Learning and Inference)：将学习与推理过程紧密结合，通过端到端可微推理和动态多源知识推理，提升系统的适应性和灵活性。

逻辑与推理 (Logic and Reasoning)：将基于逻辑的推理方法与神经网络深度融合，涵盖逻辑推理、概率推理以及神经符号系统的语法和语义分析。

可解释性与可信度 (Explainability and Trustworthiness)：构建可解释的模型和推理机制，确保神经符号系统的可信度和可靠性，增强用户对系统的信任。

元认知 (Meta-Cognition)：通过整合神经网络与符号表示，使系统能够实现自我监测、评估和调整推理与学习过程，提升系统的自适应能力。

2.1. 知识表示

在知识表示的研究领域，当前的重点在于**推进语义接地 (grounding)**、**复杂关系的再现**以及**数据效率的提升**。近年来，常识知识库和基于事件的表征方法的发展显著提升了人工智能对日常事件的理解能力，并有效降低了文本生成中的错误率。这些研究进一步深化，通过探索少样本学习模型中常识知识的最小数据需求，以及利用神经符号表示来提高训练效率并降低成本，取得了显著进展。

此外，通过预测复杂关系和知识图谱中的嵌入技术，知识表示的精细化得到了显著提升。同时，个性化知识的整合被证明能够确保叙事主体在讲故事过程中的连贯性和一致性。例如，NeuroQL——一种用于主体间推理的领域特定语言，通过捕捉复杂和远程的关系，展示了神经符号方法在显著节省训练时间和环境影响方面的巨大潜力⁴。

然而，尽管取得了诸多突破，神经符号人工智能在知识表示领域的研究仍面临一些亟待解决的科学问题。如何增强符号的动态解释和操纵能力，发展元认知能力以监测和调整推理过程，以及确保推理路径的透明性和可解释性，使其更接近人类的思维方式、更具适应性且更强大，仍是当前研究的重要方向。

⁴ N. Papoulias, Neuroql: A neuro-symbolic language and dataset for inter-subjective reasoning, arXiv preprint arXiv:2303.07146 (2023). URL: <https://github.com/orgdlabs/neuroQL>. doi:10.1109/cict56698.2022.9997814, submitted 13 March, 2023; originally announced March 2023.

2.2.学习和推断

在学习与推断领域，研究主要聚焦于**面向增强学习的神经符号集成**、**高级问题解决与决策能力的提升**，以及**用于增强模型可信度的语义强化**。通过将符号推理与神经学习机制有机结合，神经符号集成在增强学习方面取得了显著进展。例如，逻辑神经网络的应用使得常识知识能够适应少样本学习场景，并将观察结果转化为逻辑事实，从而为复杂任务提供更有效的解决方案。

在高级问题解决与决策方面，Plan-SOFAI⁵和 ZeroC 架构⁶等研究展示了神经符号方法的强大潜力。这些方法通过增强人工智能规划和零样本概念识别能力，成功集成了快速和慢速思维模型，显著提升了机器的泛化能力，使其能够在复杂多变的环境中做出更精准的决策。

语义增强和模型可信度的提升也是当前研究的重要方向。通过引入自回归模型的伪语义损失，将逻辑约束融入损失函数，并利用逻辑张量网络将逻辑与神经网络相结合。这种创新方法不仅提高了模型的逻辑一致性，还降低了模型毒性，提升了预测准确性，为构建更可靠的人工智能模型奠定了基础。

然而，尽管在这些领域取得了诸多突破，神经符号人工智能在学习和推断领域的研究仍面临一些亟待解决的科学问题。未来的研究需要开发增量学习机制，使符号系统能够随着新体验的积累而动态进化；创建基于情境线索的情境感知推理机制，以适应复杂多变的环境；实现复杂推理链的细粒度可解释性，使推理过程更加透明；并探索元认知能力，使系统能够在动态环境中自主监控、评估和优化其学习过程。这些研究方向将为神经符号人工智能的进一步发展提供新的动力和机遇。

2.3.逻辑与推理

在神经符号逻辑与推理的研究领域，当前的研究重点主要集中在**逻辑推理与概率模型的融合**、**常识与语言理解的提升**以及**决策能力的增强**。例如，逻辑信用

⁵ F. Fabiano, V. Pallagani, M. B. Ganapini, L. Horesh, A. Loreggia, K. Murugesan, F. Rossi, B. Srivastava, Plan-SOFAI: A neuro- symbolic planning architecture, in: Neuro-Symbolic Learning and Reasoning in the era of Large Language Models, 2023. URL: <https://github.com/FrancescoFabiano/SOFAI-x-Planning>. arXiv:<https://openreview.net/forum?id=ORAhay0H4x>.

⁶ T. Wu, M. Tjandrasuwita, Z. Wu, X. Yang, K. Liu, R. Sosič, J. Leskovec, Zeroc: A neuro-symbolic model for zero-shot concept recognition and acquisition at inference time, arXiv preprint arXiv:2206.15049 (2022). URL: <https://github.com/snap-stanford/zeroc>. doi:10.1109/tmm.2023.3318300, submitted 11 October, 2022; v1 submitted 30 June, 2022; originally announced June 2022.

网络（Logic Credit Network）通过将逻辑推理与概率模型相结合，有效处理不精确信息；DeepStochLog 则利用神经网络增强传统逻辑编程，以应对复杂的推理任务⁷。

在常识性知识与语言理解方面，相关研究取得了显著进展。例如，Kogito 能够从文本输入中生成常识性知识推断，从而增强人工智能系统的适应性⁸；LinkBERT 则通过整合文档链接，显著提升了语言理解和推理能力，尤其在多跳推理任务（Multi-hop reasoning tasks）中表现出色⁹。

为了增强决策能力，相关研究提出了多种创新方法。例如，神经符号常识社会推理将神经符号方法集成到自主系统中，以提升其决策能力¹⁰；LASER 则巧妙地结合了神经网络的灵活性与符号逻辑的精确性¹¹。此外，针对大型语言模型（LLM）在可信度和推理方面的局限性，通过生成式人工智能与符号人工智能的系统集成，以提高人工智能系统的可靠性。

然而，尽管取得了诸多进展，神经符号人工智能在逻辑与推理领域的研究仍面临一些亟待解决的科学问题。未来的研究需要进一步开发可扩展的框架，将传统逻辑编程与神经网络有机结合，以应对复杂的推理任务。同时，研究还需深化常识知识与高级语言理解的结合，进一步提升多跳推理能力。此外，如何更好地整合符号逻辑与神经网络，以确保决策的可靠性和可信度，也是当前研究的重点。最后，引入元认知能力，实现推理策略的自我监控与调整，从而获得更清晰、更易于理解的解释，也是未来研究的重要方向。

2.4.可解释性和可信度

在神经符号人工智能领域，围绕可解释性和可信度的研究重点在于**推进自然语言处理（NLP）技术、增强逻辑推理能力以及完善语言理解和总结**。例如，通过引入了一种带有概率规则的逻辑推理器，融合符号和神经知识来解决脆性匹配

⁷ T. Winters, G. Marra, R. Manhaeve, L. De Raedt, Deepstochlog: Neural stochastic logic programming, 2021. URL: <https://github.com/ML-KULeuven/deepstochlog>. doi:10.48550/ARXIV.2106.12574.

⁸ M. Ismayilzada, A. Bosselut, kogito: A commonsense knowledge inference toolkit, 2022. URL: <https://github.com/epfl-nlp/kogito>. doi:10.48550/ARXIV.2211.08451.

⁹ M. Yasunaga, J. Leskovec, P. Liang, Linkbert: Pretraining language models with document links, 2022. URL: <https://github.com/michiyasunaga/LinkBERT>. doi:10.48550/ARXIV.2203.15827.

¹⁰ D. Chanin, A. Hunter, Neuro-symbolic commonsense social reasoning, arXiv preprint arXiv:2303.08264 (2023). URL: <https://github.com/chanind/amr-logic-converter>, submitted 14 March, 2023; originally announced March 2023.

¹¹ J. Huang, Z. Li, M. Naik, S.-N. Lim, Laser: A neuro-symbolic framework for learning spatial-temporal scene graphs with weak supervision, arXiv preprint arXiv:2304.07647 (2023). URL: <https://github.com/isLinXu/paper-list>. doi:10.23919/eusipco58844.2023.10289877, submitted 22 November, 2023; v1 submitted 15 April, 2023; originally announced April 2023.

问题，从而显著提升了逻辑推理的鲁棒性和准确性。此外，结构感知抽象对话通过整合话语关系、动作三元组和结构化图，实现了精确且富有上下文的摘要生成，推动了 NLP 技术的进步。

在语义层面，通过识别和纠正神经符号场景中的“混杂因素”，提高人工智能决策的清晰度，增强了逻辑推理的可理解性，从而培养了用户对系统的信任。例如，FactPEGASUS 专注于通过优化预训练和微调方法来确保摘要的真实性，这对于保持摘要的完整性和提升语言理解的精炼性至关重要¹²。此外，神经符号方法通过逻辑推理和线索检测，增强了可解释的简短答案评分，将人工智能能力与类人反应联系起来，进一步推动了自然语言处理技术的发展。

尽管取得了诸多进展，神经符号人工智能在可解释性和可信度领域的研究仍面临一些亟待解决的科学问题。未来的研究需要探索如何使符号表示能够实时适应和进化，以保持系统的透明度；如何整合元认知机制，实现自我监控和调整推理策略；如何为复杂的认知任务开发更具解释性的 NLP 技术；以及如何确保人工智能输出的事实一致性，同时为底层推理过程提供清晰、详细的解释。这些问题的解决将为神经符号人工智能在自然语言处理领域的进一步发展奠定坚实基础。

2.5.元认知

元认知是指个体对自身思维过程的思考、监控和调节，是一种在认知任务中实现自我意识和自我调控的高级认知能力。它在神经符号人工智能中扮演着“认知任务的总指挥”角色，能够高效地引导系统资源，精准分配到处处理任务的合适模块中。这种能力对于需要反思、规划和适应的任务至关重要，它能够显著增强学习、问题解决和决策的能力，因此成为神经符号人工智能的关键研究方向之一。然而，目前神经符号人工智能领域的研究尚未充分涵盖元认知。忽视元认知限制了系统的自主性、适应性和可靠性，阻碍了其在动态环境中进行有效纠错和提升可信度的能力。因此，将元认知融入神经符号人工智能系统，使其具备自我意识和自我调节能力，将成为未来进步的关键。

¹² D. Wan, M. Bansal, Factpegasus: Factuality-aware pre-training and fine-tuning for abstractive summarization, 2022. URL: <https://github.com/meetdavidwan/factpegasus>. doi:10.48550/ARXIV.2205.07830.

近年来，元认知领域的研究取得了显著进展。例如，通过强化学习（RL）来近似元认知能力，将认知架构与大型语言模型（LLM）集成以模拟元认知功能，以及开发多种人工智能架构和系统以展示认知通用模型（CMC）的可行性。这些研究展示了将符号特征与 RL 算法集成的益处，还探索了通过将认知结构与 LLM 融合来增强通用智能，创建利用符号和亚符号人工智能方法优势的具身代理。此外，通过模块化、代理和神经符号方法，将 LLM 与认知架构相结合，开发出更稳健的人工智能系统。这些项目与认知通用模型对齐，整合了像 ACT-R、Soar 和 Sigma 这样的认知架构，为人类认知提供了一个统一的框架。尽管已有理论提出通过神经符号方法，将生成式网络与认知通用模型结合以模拟人类认知过程，以实现强大且可解释的人工智能系统，但这一目标尚未完全实现。

尽管取得了诸多进展，神经符号人工智能在元认知领域的研究仍面临一些亟待解决的科学问题。例如，如何将符号推理与元强化学习相结合以实现复杂决策；如何将认知架构与 LLM 融合以开发具有元认知能力的代理；如何利用 LLM 通过元认知信号增强基于实例的学习；如何为实时冲突解决创建自适应的元认知框架；如何将模块化和代理方法相结合，构建与认知通用模型一致的元认知人工智能系统；如何用元认知意识的认知架构改进少样本学习，以及如何开发能够复制类人元认知过程的神经符号生成网络。这些问题的解决将为神经符号人工智能的发展提供新的方向和机遇。

三、神经符号 AI 的基础研究方向

AlphaGeometry 是一种针对奥林匹克竞赛级别的平面几何问题的神经符号系统。它通过在大规模合成数据上训练的神经语言模型来指导符号推理引擎，能够高效地合成数百万个定理和证明¹³。作为一项标志性研究，AlphaGeometry 是知识表示、学习与推断、逻辑与推理、可解释性与可信度等多个研究领域的深度融合。它展示了神经符号人工智能在高级问题解决能力方面的巨大潜力，弥合了人工智能研究多个领域的差距，表明该领域在跨学科研究方面具有巨大的发展空间。

尽管将神经符号人工智能初步表述为系统 1（快速、直觉化）和系统 2（慢速、理性化）的思维方式是一种有用的工具，有助于将该领域推向神经与符号过

¹³ C. Yang, S. Chaudhuri, Safe neurosymbolic learning with differentiable symbolic execution, arXiv preprint arXiv:2203.07671 (2022). URL: <https://github.com/cxyang1997/dse>. doi:10.23919/fmcad.2018.8602991, submitted 15 March, 2022; originally announced March 2022.

程整合的共同方向，但目前对人类认知处理能力的近似仍然过于简化，尚未完全捕捉到推动该领域发展的系统级分解的全貌。正如丹尼尔·卡尼曼所指出的：“这两个系统在大脑中或任何其他地方并不存在。‘系统 1 做 X’是一种简化的说法，意思是‘X 是自动发生的’；而‘系统 2 被调动去做 Y’也是一种简化的说法，意思是‘警觉性增强，瞳孔放大，注意力集中，并且执行了活动 Y’。”人类水平的认知能力实际上源自一个极其复杂、层次丰富且紧密相连的系统之系统，这些系统相互统一，同时表现出系统 1 和系统 2 的特征。

神经符号人工智能研究的目的是创建一个具有强大推理能力的高级混合人工智能模型。因此，为了推动该领域朝着这一目标进一步发展，必须构建能够像丹尼尔·卡尼曼的系统 1 和系统 2 那样运作的体系，通过元认知进行调控，能够在必要时表现出高效且“懒惰”的行为，又在需要时表现出专注且精确的行为。这种系统不仅能够更好地模拟人类的认知过程，还将为人工智能的发展开辟新的可能性。

神经符号AI，赋能绿色制造的人工智能引擎
<https://www.nsaihome.org.cn>



NSAiHome
神经符号人工智能社区

（责任编辑 曹晓舟，审核 刘永光）