arXiv:2611.28356v2 [cs.AI] 24 Nov 2026

Engram：异构模型智能体调优的多阶段记忆

Huanran Li、Nhat X.T. Le、Kevin De Angeli、Yao Xuan、Alireza Vahdatpour、Djordje Gligorijevic

Meta Platforms, Inc.

如今，大语言模型（LLM）智能体已经能够调优神经网络，但由于缺乏跨会话记忆，它们往往要耗费高昂的试验成本来重新发现已知经验。由于生产模型在架构、训练设置和推理预算方面各不相同，经验复用十分困难。我们提出了 Engram，这是一种记忆系统，可将过往试验结果与其背后的假设和理由关联起来，并评估这些证据是否适用于当前模型。在线代理会在需要证据的决策点整理相关信息：初始指导在开始时提供经验证的经验；会话检索在生成每一批假设之前提供相关的过往会话；试验候选项审查则会将每个试验候选项与匹配的过往试验进行核对。在会话之间，离线智能体会提炼经验证的经验，使记忆系统能够不断自我改进。在两个广告模型上，Engram 仅使用少 47% 的试验次数，就达到了无记忆对照组经过 32 次试验所取得的结果，并且最终平均获得了更优的模型。

日期： 2026年11月**通讯：** HuanranLi，邮箱：huanranli@meta.com

Meta

1 引言

深度神经网络的调优长期以来依赖人类专业知识：工程师根据经验和文献选择实验，然后检查结果。贝叶斯优化（Snoek 等，2012）减少了这项工作，但其搜索的是一个固定的、主要由数值构成的空间，并且必须预先指定。大语言模型（LLM）智能体通过直接对模型及其配置进行推理，消除了这一限制。给定当前的基线配置、目标和计算预算后，智能体会读取模型代码和日志，提出带有理由的假设，运行一次训练试验，并利用结果选择下一项实验（Yao 等，2023；Yang 等，2024）。此类智能体已证明其有效性：在 16 个表格数据 Kaggle 任务中，AIDE 的平均表现优于 H2O AutoML （ Jiang 等，2025 ），而面向推荐系统的智能体报告称，其已投入生产，并在不损失业务指标的情况下将训练时间缩短了八倍（Wang 等，2026）。

随着智能体逐渐成为模型调优的工具，它们不可避免地需要对许多异构模型进行调优，而且每个模型都要反复调优。例如，大型广告平台维护着一个庞大的生态系统
这些模型服务于不同的产品界面、排序阶段和广告主目标（Liu 等，2026）。生产模型会定期使用新数据重新训练、扩展新特征，并围绕新架构重新设计，而每项变更都需要开启新的调优会话。然而，由于生产模型在目标、特征和架构，训练设置（数据、运行时长和优化器调度），以及训练数据的时间跨度上各不相同，因此很难复用先前的会话。根据我们调优生产模型的经验，我们发现了异构模型智能体记忆面临的五项挑战。**（C1）以往的收益取决于具体模型和训练设置：**更大的嵌入维度等变更所产生的效果会因模型、其基线和训练设置而异，因此在一个模型上取得的收益未必能在另一个模型上重现。**（C2）变更是否适用取决于推理预算：**有助于浮点运算次数预算充裕的模型的变更，在预算紧张时可能不适用。**（C3）仅凭结果无法解释结果：**配置和结果记录通常无法揭示结果成立所依据的假设，因此无法告知后续智能体该结果是否适用。**（C4）当前调优会话期间，过往会话的相关性会发生变化：**一旦当前模型的结果表明哪些变更有效，最有用的过往会话便是那些从相同变更中获益的会话，而非启动时最相似或时间上最近的会话。**（C5）错误套用证据会使搜索效果变差：**来自与当前模型差异过大的模型的过往结果，可能会将智能体引向更差的配置

。Engram 通过结合当前模型和预算来解读过往试验结果及其理由，并在决策阶段提供适用证据以引导搜索，从而应对这一问题（图 1）。

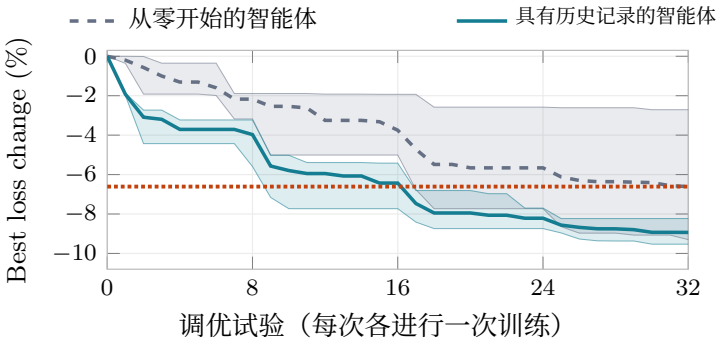


图1 广告模型调优：未见数据上的最佳归一化对数损失（五个会话的均值和范围）。借助历史记录，智能体只需少进行47%的试验，即可达到其从零开始时进行32次试验所取得的结果。

相关工作。以往关于在大语言模型智能体中复用经验的研究主要分为三类。用于模型调优的大语言模型智能体会在任务内部复用较早的解决方案和结果（Jiang et al., 2025），提供按得分排序的过往配置日志（Wang et al., 2026），或在不同数据集之间传递提炼出的启发式规则（Cao et al., 2026a），但都没有判断某个结果是否适用于另一个模型或预算（C1、C2、C5）。我们采用一份按性能增益排序的过往会话日志作为基线，以衡量加入此类判断后带来的价值。经验记忆系统提供提炼出的洞见和相似轨迹（Zhao et al., 2024），对检索到的经验进行重新排序和改写（Cao et al., 2026b），或学习哪些记忆有用（Zhang et al., 2026）以及何时停止检索（Xia et al., 2026），但既不进行相对于基线的比较，也不检查预算。贝叶斯优化的迁移方法会从相似的过往数据集开始搜索（Feurer et al., 2015），或根据学习到的相似度为过往任务赋予权重（Perrone et al., 2018），但只迁移数值结果，没有提供能够解释某个结果是否应当迁移的理由（C3）。

图 2 概述了 Engram。我们的贡献如下：

- **试验结果与理由的联合分析：** 关系型存储将每次过往试验的结果与其假设和理由关联起来，基于大语言模型的在线代理利用这两者判断检索到的结果是否有充分依据，以及是否适用于当前模型和预算（C1、C3）。
- **多阶段检索：** 初始指导（IG）在启动时提供范围与当前模型匹配的经验证经验，会话检索（SR）在每个批次之前检索相关的过往会话，并根据当前反馈对其重新排序，而试验候选项审查（TCR）则根据匹配的过往试验和预算检查每个试验候选项（C2、C4、C5）。
- **通过知识整合实现自我改进的记忆：** 在会话之间，离线智能体从已完成的会话中提炼经验，在后续会话中对其进行验证，并将其应用于之后的会话。
- **初步证据：** 在搜索早期，Engram 在两个广告模型上优于无记忆对照组；对于历史表现不佳的模型，Engram 也优于按性能排序的日志。

2 Engram设计

2.1 问题设定

调优会话从基线配置 x_0 开始，在固定的评估环境中进行，最多可提交 K 次试验，每次试验评估一组搜索参数 x_k 配置，例如学习率和层宽度。每次试验都会测量质量指标 $q(x)$（数值越低越好）以及推理成本 $c(x)$；如果试验的成本不超过基线成本的 ρ 倍，则该试验可行。智能体每次只能依据

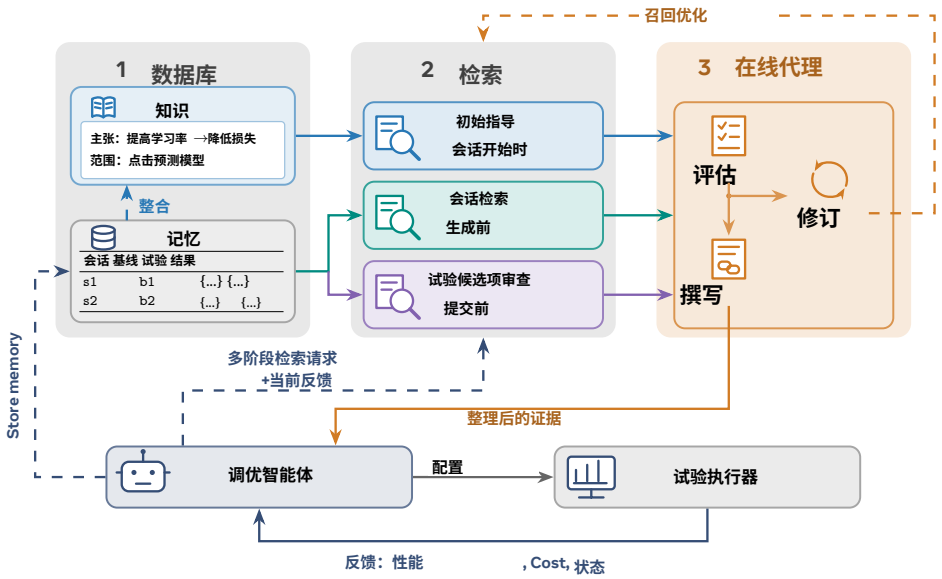


图2 在线智能体： 为调优智能体进行多阶段检索与证据整理。插图为示意图。

截至目前已完成试验的结果，而会话返回

$$x^\star = \underset{x\in\{x_0,\dots,x_K\}}{\arg\min} \quad q(x) \quad \text{s.t.} \quad c(x) \le \rho\, c(x_0). \tag{1}$$

在我们的设定中， q 表示归一化熵（NE），即对数损失除以平均标签率的熵（He 等人，2014），而 c 表示推理浮点运算次数；下文沿用这些名称。调优智能体以大语言模型作为优化器（Yang 等人，2024），运行观察–推理–行动循环（Yao 等人，2023）：它读取包含通用建议的调优指南和当前会话的结果，记录每个假设及其理由，并提交试验。Engram 通过提供过往会话中的证据补充这一反馈，而这些证据原本不会为智能体所见。

2.2 记忆、知识与整合

Engram 维护两个存储库，分别存放原始证据和经验证的经验。记忆保存原始证据，每日回填会从每个已结束的会话中摄取这些证据。对于每次试验，它都会记录配置变更、NE、浮点运算次数和状态，以及调优智能体的假设、理由和分析。理由在试验运行前撰写，用于解释为何该假设应能改善 NE；分析则在试验结束后解读测得的结果。知识则保存离线整合从记忆中提取并验证的经验。每条经验都指定一个参数、一组初始值范围，以及在该范围内应采取的调整操作；三者共同定义一次调整操作。它还会说明关于该调整操作对 NE 影响的主张，并界定该主张适用的模型范围，例如具有特定架构或使用特定数据量训练的模型。

离线整合 此过程从所有模型中提炼证据，形成全局经验，并且只有在验证每条经验后，才将其存入知识。这样一来，离线智能体便将记忆 转化为知识，使 Engram 成为一种自 我改进的记忆 ，其指导能力会随着调优模型数量的增加而增强。代码首先从会话内的试验对中挖掘候选经验，衡量每项调整操作相对于其所修改的配置对 NE 造成的变化。单参数变更和联合变更分开挖掘；对于联合变更所得的经验，只将其表述为相关性，因为无法将其效果归因于某一个参数。随后，离线智能体会结合每条候选经验的支持试验和记录的理由进行审查，决定保留、缩小其适用范围或将其舍弃，并将其与调优指南中的默认建议联系起来。只有满足以下条件的候选经验才会发布：(1) 无论是在完整历史数据上还是在时间交叉验证中，其结果预测准确度都高于默认调优指南；(2) 其依据来自多个不同模型（默认五个）；(3) 获得多个独立大语言模型评审的认可（默认两个）。

算法 1 Engram 在一次调优会话中的应用

要求： 基线配置 x_0，试验预算 K，浮点运算次数上限 ρ

编写初始指导（*IG*）简介
对于预算内的每一批试验 K **，执行**
    对目前观察到的结果运行会话检索（*SR*）
    重复
        调优智能体提出试验或修改被标记的试验
        对提出的试验执行试验候选项审查（*TCR*）
    直到没有试验被标记或达到审查上限
    执行试验并记录其结果
结束循环
将会话存储到记忆中，以便进行离线整合
返回 $x^\star$，最佳可行配置

2.3 多阶段检索

算法 1 展示了各阶段何时运行：会话开始时运行一次初始指导，随后会话按轮次推进，直至耗尽试验预算。每一轮中，会话检索会根据目前为止的结果检索过往会话，调优智能体提出一批试验，试验候选项审查则会在提交前对其进行检查。在每个阶段，基于大语言模型的在线代理都会结合当前基线和结果接收检索到的证据，并在循环中运行：评估这些证据是否适用于当前模型；若证据不足，则修订检索结果；并编写调优智能体读取的文本。接下来，我们将介绍各阶段如何检索和整理证据。

初始指导（*IG*）会话开始时，初始指导会告知调优智能体哪些经验证的经验可应用于其模型。若最新一次整合运行中的某条经验满足以下条件，则可应用于当前模型：**(1)** 其范围涵盖当前模型；**(2)** 模型支持该经验涉及的参数；**(3)** 该参数的基线值处于调整操作的起始范围内。在线代理会根据基线配置审查所有适用的经验，并撰写一份简报，推荐这些经验所支持的调整操作，同时将联合变更经验作为相关性而非建议进行报告。

会话检索（*SR*）在调优智能体生成每一批假设之前，会话检索会检索与当前会话相似的过往会话，并说明这些会话中的发现。它根据加权元数据匹配度，以及假设与结果之间的加权一致性，对每个过往会话 c 进行排名：

$$\mathrm{score}(c) = \sum_i w_i \,\mathrm{Match}_i(c) + \lambda \,\mathrm{Agreement}(c). \tag{2}$$

每个匹配项$_i(c) \in [0,1]$ 会比较 c 的一个元数据字段与当前会话中的对应字段，例如模型架构、模型规模或已探索参数。权重 w_i 用于设定各字段的重要性。一致性(c)衡量当前会话迄今为止的 NE 增益中，有多少来自这样一些假设：它们调整了 c 也曾探索并取得改进的参数。权重 λ 用于设定其相对于元数据匹配的重要性。由于每一批假设完成后都会重新计算一致性，排名也会随着新结果不断变化。在线代理会收到排名后的会话及其测量结果和理由，并检查这些会话是否涵盖当前模型可调优的参数、是否符合 FLOPs 预算，以及是否足够近期；它还可以排除某些会话，或调整 w_i 和 λ 以检索修订后的会话集。随后，它会撰写建议，引用保留会话的测量结果和理由，而下一项假设的选择仍由调优智能体决定。

试验候选项审查（*TCR*）调优智能体提出一批试验后，试验候选项审查会针对做出相同调整的过往试验，逐一检查每个试验候选项在记忆中的记录。会话检索会对完整会话进行推理，而试验候选项审查则会检查与拟议调整相关的单个试验。在线代理会收到每个候选项，以及调整了相同参数的过往试验。它会判断该候选项是否可能降低归一化熵（NE）并符合 FLOPs 上限；对于任何被判定为不太可能降低 NE 或超出上限的候选项，调优智能体都必须将其替换或说明理由。

表 1 实验臂在 8、16、24 和 32 次试验后，五个会话中的平均最佳可行归一化熵变化（%）。粗体：领先于对照（例如精确 Mann–Whitney U，$p < 0.05$，未校正）。

	模型 1				模型 2			
Arm	8	16	24	32	8	16	24	32
基线								
对照（无记忆）	−2.17	−3.75	−5.66	−6.61	−0.29	−0.63	−0.85	−1.11
按 NE 排序的日志	**−3.55**	**−6.01**	−7.91	−9.09	−0.32	−0.65	−0.96	−1.26
Engram 阶段								
仅初始引导	−2.97	−5.88	−7.62	−8.66	**−0.48**	−0.78	−0.95	−1.15
仅会话检索	**−4.17**	**−6.48**	−7.97	−9.01	−0.32	−0.61	−1.11	−1.34
初始引导 + 会话检索	**−3.47**	**−5.85**	−7.37	−9.03	**−0.44**	**−0.95**	−1.18	−1.35
完整 Engram	**−3.97**	**−6.43**	**−8.22**	−8.93	−0.40	**−0.96**	**−1.33**	**−1.46**
完整 *Engram* 的消融实验								
不进行筛选	**−3.49**	−5.89	−7.66	−8.63	−0.41	−0.65	−1.10	−1.33
不达成一致	**−4.04**	−5.95	−7.42	−8.76	**−0.48**	−0.89	−1.10	−1.27
仅使用远期历史记录	**−4.79**	**−6.40**	−7.59	−8.65	与完整 Engram 相同			

3 实验

我们研究历史证据是否有助于智能体更早找到有用的调优假设，以及哪个阶段提供了关键证据。如果某个过往模型与当前模型属于同一种生产模型类型，我们就称其相近，否则称其相远。模型类型决定广告产品、排序阶段、预测事件和归因窗口，因此也决定其服务的流量；相近的模型在架构、特征和训练方案上仍有所不同。我们研究了两个历史记录不同的广告模型。模型 1 是主要研究对象，拥有丰富的历史记录，共有 227 个会话，其中 59 个来自相近模型，另有 8 个基线。模型 2 用于检验研究发现能否推广到历史记录匮乏的情形：其 299 个会话全部来自相远的模型，不过其中一些模型与它属于同一架构系列。两者的历史记录均不包含当前评估基线。表格1总结了各实验组的设计。

3.1 实验设置与指标

两个模型均使用八个实验组进行评估：无记忆对照组、用按 NE 排序的日志替代 Engram、仅使用初始指导、仅使用会话检索、同时使用初始指导和会话检索、完整 Engram，以及不进行在线整理或不具备反馈一致性的完整 Engram。模型 1 还额外运行了仅使用远距离历史的完整 Engram；对于模型 2，由于其全部历史记录均属于远距离历史，该实验组与完整 Engram 重合。每个实验组包含五个会话，在推理浮点运算次数上限为 $\rho = 1.5$ 的条件下，以同步的八个一批提交试验。每个试验在模型 1 的约 1.9 亿个样本和模型 2 的约 1.44 亿个样本上训练一轮次，每批分别耗时约五小时和三小时。除批大小、轮次数和浮点运算次数上限外，搜索空间不受限制；智能体主要调整学习率（LR）、预热计划、优化器设置、嵌入和层大小，以及输入特征。调优智能体及其子智能体使用 Claude Opus 5.5，在线代理则使用规模较小的 Claude Sonnet 5.5，以降低 Engram 的成本和延迟，因为在线代理的每次调用都会在不使用工具的情况下评估所提供的证据。在每个会话中，初始指导运行一次，会话检索最多检索 50 个会话，并且每批最多修订两次检索结果；试验候选项审查则对每批检查一次。实际上，这些模型上的调优会话的试验次数中位数为 9–10 次，90% 的会话在 25 次以内结束，因此早期收益最为重要；我们分析了 32 次试验，以验证这些收益能够持续。

指标。 由于每次试验只训练一个轮次，并且每个样本都会在模型用其训练之前进行评分，因此归一化熵始终是在未见数据上测量的。我们将归一化熵报告为相对于基线的百分比变化，因此负值表示有所改善。我们报告每个实验组在 8、16、24 和 32 次试验后可行的最佳归一化熵变化均值，这既反映搜索速度，也反映最终质量，差异以百分点表示。我们对两个实验组各自五个会话的结果进行双侧精确曼–惠特尼U检验。我们采用秩检验，因为五个会话不足以支持正态性假设，而且结果通常呈双峰分布：会话要么找到关键变化，要么陷入停滞。

3.2 速度与最终质量

Engram 在两个模型上都能更早达到较强的配置；在模型 2 上，它还取得了更好的最终结果（表 1）。在两个模型上，完整 Engram 的均值都以少 47% 的试验次数（17 次而非 32 次）达到对照组 32 次试验后的均值。在模型 1 上，完整 Engram、仅会话检索、初始引导加会话检索、按 NE 排序的日志以及仅使用远距离数据的实验组，在 8 次和 16 次试验后均领先于对照组（$p \leq 0.008$），而完整 Engram 在 24 次试验后仍保持领先。此后，对照组逐渐缩小差距：其中三个会话各自独立发现了关键的学习率杠杆（第 3.3 节），并达到 -8.89% 至 -9.30%；另两个会话则始终未能发现该杠杆，结果维持在约 -3%。在模型 2 上，单独使用任何一个阶段都不足以取得优势。仅初始引导只在 8 次试验后领先于对照组，并直到第 30 次试验才达到对照组 32 次试验后的均值；而仅会话检索在前 16 次试验中与对照组持平。相比之下，完整 Engram 在 16 次和 24 次试验后领先于对照组（$p = 0.008$），在 32 次试验后仍保持领先（$p = 0.032$；-1.46% 对比-1.11%），并在 16 次和 24 次试验后领先于两个单独阶段（$p = 0.008$）。按 NE 排序的日志揭示了 Engram 的适用性判断何时发挥作用。在模型 1 上，它在每个检查点的表现都与完整 Engram 持平（$p \geq 0.22$），因为这里的 59 个过往会话来自相近模型。在模型 2 上，整个历史记录都来自远距离模型，因此它从未显著领先于对照组，并且在 16、24 和 32 次试验后均落后于完整 Engram（$p \leq 0.032$）。

3.3 证据如何到达搜索过程

决定性证据在两个模型上分别源自不同阶段，这支持多阶段检索。对于每个决定性假设，表2报告了提供证据的阶段、用于检验该假设的对照会话和完整 Engram 会话的数量，以及证据出现的时间。这项回溯性分析并未分离各阶段的单独影响。在模型 1 上，初始指导提供了一条经验：将非嵌入参数的学习率（三密集学习率）从基线值 0.04 提高到 0.08–0.02 调优指南范围之外的区域，即 0.04 的三倍，并进入指南警告会发散的区域。每个完整 Engram 会话都将学习率提高到 0.12 或更高（首次出现在第 10–25 次试验），而对照会话中只有三个这样做（出现在第 18–29 次试验），并且从未这样做的两个会话最后完成。然而，仅凭初始引导，在任何检查点上相对于对照的领先都不显著（$p \geq 0.056$）。会话检索还从相近模型的会话中提供了适当的嵌入大小：每个完整 Engram 会话都在第一次试验中采用该大小，而对照组和仅远程模型组要到第 2–7 次试验才找到该大小，并且分别花费 21 次和 11 次试验使用更大的大小。在模型 2 上，初始学习率经验只是将第一次提高学习率的时间提前到第 1 次试验；决定性的变化则是缩短密集学习率和稀疏学习率的预热时间。该调度方案只出现在远程模型的历史记录中：每个完整 Engram 会话都在第 10–15 次试验采用该方案，并引用一个被检索会话作为来源；而三个对照会话从第 18 次试验起独立发现该方案，会话检索单独使用时则要到第 13–26 次试验才首次采用。

表2 32 次试验内的决定性假设：阶段、检验该假设的会话数、首次试验中位数（>32 若从未检验）以及对照 / Engram 的 NE 平均变化（仅计入得出该假设的试验）。粗体表示 Engram 领先。

模型	假设	阶段	会话	首次试验	归一化熵 (%)
1	密集学习率 $0.04 \to \geq 0.12$	IG	3 / **5**	29 / **18**	-8.29
1	嵌入大小 $4 \to 8$	SR	5 / 5	4 / **1**	-5.74
2	密集学习率 > 0.04	IG	5 / 5	6 / **1**	-0.73
2	较短的学习率预热	SR	3 / **5**	28 / **11**	-1.17

4 结论

Engram 将测量结果与其理由关联起来，在调优的三个阶段整理证据，并在线下验证经验。在两个广告模型上，它仅用少 47% 的试验次数，便达到无记忆对照组经过 32 次试验取得的结果；其中一个模型的历史记录完全来自较早的模型，而且两个模型上的关键证据分别源自不同阶段。这些初步发现仅涵盖两个模型、每组五个会话；更广泛的迁移效果，以及自我改进记忆在反复整合过程中的益处，仍有待评估。Engram 的核心设计原则是评估哪些过往证据适用，并在调优智能体需要时提供这些证据，而非仅仅提供更多历史记录。

伦理声明

Engram 仅读取已完成调优会话的配置、指标和理由，不收集任何用户数据。

参考文献

Shiteng Cao、Jingwen Liu、Junda She 和 Zhiheng Li。VirtualMLE：优化序列推荐系统的虚拟机器学习工程师。arXiv:2606.03221，2026a。Zouying Cao、Jiaji Deng、Li Yu 等。Remember Me, Refine Me：用于经验驱动智能体演进的动态程序记忆框架。发表于*ACL Findings*，2026b。Matthias Feurer、Jost Tobias Springenberg 和 Frank Hutter。通过元学习初始化贝叶斯超参数优化。发表于$AAAI$，2015。Xinran He、Junfeng Pan、Ou Jin 等。在 Facebook 预测广告点击的实践经验。发表于$ADKDD$，2014。Zhengyao Jiang、Dominik Schmidt、Dhruv Srikanth 等。AIDE：在代码空间中进行人工智能驱动的探索。arXiv:2502.13138，2025。Jiang Liu、John Martabano Landy、Yao Xuan 等。一次设计，规模化部署：面向大模型生态系统的模板驱动机器学习开发。arXiv:2603.24963，2026。Valerio Perrone、Rodolphe Jenatton、Matthias W. Seeger 和 Cédric Archambeau。可扩展的超参数迁移学习。发表于 $NeurIPS$，2018。Jasper Snoek、Hugo Larochelle 和 Ryan P. Adams。机器学习算法的实用贝叶斯优化。发表于 $NeurIPS$，2012。Haochen Wang、Yi Wu、Daryl Chang、Li Wei 和 Lukasz Heldt。自进化推荐系统：借助大语言模型智能体实现端到端自主模型优化。发表于 $RecSys$，2026。Menglin Xia、Xuchao Zhang、Shantanu Dixit 等。Memora：一种平衡抽象与特异性的谐和记忆表示。发表于$ICML$，2026。Chengrun Yang、Xuezhi Wang、Yifeng Lu 等。大语言模型即优化器。发表于 $ICLR$，2024。Shunyu Yao、Jeffrey Zhao、Dian Yu 等。ReAct：协同语言模型中的推理与行动。发表于 $ICLR$，2023。Shengtao Zhang、Jiaqian Wang、Ruiwen Zhou 等。MemRL：通过对情景记忆进行运行时强化学习实现自进化智能体。arXiv:2601.03192v2，2026。Andrew Zhao、Daniel Huang、Quentin Xu 等。ExpeL：大语言模型智能体是经验型学习者。发表于$AAAI$，2024。