物有所值：使用重置与丢弃（ReD），在固定预算下提升大语言模型的推理能力

Sagi Meir [1 2] **Tommer D. Keidar** [1 2] **Noam Levi** [3] **Shlomi Reuveni** [1 2 4] **Barak Hirshberg** [1 2 4]

arXiv:2601.21522v1 [cs.LG] 29 Jan 2026

摘要

大语言模型（LLM）在可验证任务上的性能通常用 pass@k 衡量，即在 k 次试验中至少正确回答一次问题的概率。在固定预算下，更合适的指标是 coverage@cost，即已回答的不重复问题的平均数量随尝试总次数变化的函数。我们将这两个指标联系起来，并表明，经验观察到的 pass@k 幂律行为会导致 coverage@cost 次线性增长（收益递减）。为解决这一问题，我们提出大语言模型的查询方法重置并丢弃（ReD），无论 pass@k 的形式如何，该方法都能在任何给定预算下提升 coverage@cost。此外，给定 pass@k，我们可以定量预测使用 ReD 可节省的尝试总次数。如果无法获得模型的 pass@k，ReD 还可以推断其幂律指数。在三个大语言模型上使用 HumanEval 开展的实验表明，ReD 能大幅减少达到目标覆盖率所需的尝试次数、词元数量和美元成本，同时还提供了一种高效测量推理幂律的方法。

1. 引言

生成多个相互独立的候选解，已成为提升大语言模型在可验证推理任务（如编程、数学和形式化证明搜索）上性能的标准方法。在不同模型系列和基准测试中，这种方法能带来显著提升：允许进行多次尝试时，相对较小的模型也能达到甚至超过较大模型的单样本性能（Chen 等人，2021；Brown 等，2024；Snell 等，2024；Wu 等，2024）。这些提升通常以 pass@k 报告：对于单个随机选择的实例，在 k 次生成中至少有一次成功解决该实例的概率（Kulal 等人，2019；Chen 等人，2021；First 等人，2023；Brown 等，2024；Hassid 等人，2024；Hughes 等人，2024；Chen 等人，2024a；Levi, 2025；Schaeffer 等，2025；Ehrlich 等人，2025；Kwok 等人，2025）。近期研究发现了一种规律性极强的经验模式：对于许多任务，$1-$pass@k 会随 k 按幂律衰减，且这一规律跨越多个数量级（Schaeffer 等，2025；Levi, 2025；Kazdan 等，2025）。

在实践中，推理计算通常由多个工作负载共享：例如，令牌上限固定的批量评测、必须尽可能处理更多工单的自动代码修复系统，或是在延迟和成本约束下运营的服务提供商。在这些场景中，相关目标并非 pass@k，而是在固定全局预算下已解决多少个不同问题。然而，在固定预算下为许多任务制定高效的分配策略，仍未得到充分探索。

为形式化这一目标，我们采用**coverage@cost** ——即投入总预算（以尝试次数、词元或美元衡量）后预期解决的独特问题数量。coverage@cost 是一种奖励解决更多问题，而非将不成比例的计算资源花在少数困难实例上的指标。关键在于，coverage@cost 不仅取决于模型，也取决于用于分配各问题尝试次数的策略。

一种常见的隐式分配方案是求解至完成：反复对一个问题采样，直到将其解决，然后再转向下一个问题。当每个问题的成功概率各不相同（即难度分布不均）时，这种策略会偏向困难问题，因为困难问题会消耗许多尝试次数，从而降低整个工作负载的吞吐量。在经验观察到的幂律 pass@k 行为下，$1-$pass@$k \propto k^{-\alpha}$，通常伴随 $0 < \alpha < 1$，这种影响可能极为显著。在这种情况下，我们证明，求解至完成会使 coverage@cost 呈次线性 增长，也就是说，追加预算带来的独特已解决问题数量会收益递减。

我们提出一种替代性的分配策略，**重置并丢弃（ReD）**：（1）将重置采样内联到下一个任务，

[1]特拉维夫大学化学学院，以色列特拉维夫 6997801。[2]特拉维夫大学生命系统物理与化学中心，以色列特拉维夫 6997801。[3]洛桑联邦理工学院（EPFL），瑞士洛桑。[4]特拉维夫大学计算分子与材料科学中心，以色列特拉维夫 6997801。通讯作者：Barak Hirshberg <hirshb@tauex.tau.ac.il>。

在固定次数的尝试 τ 后（极端情况下，每次尝试后），以及（2）一旦问题已解决，就立即丢弃该问题，从而不再为其投入额外计算资源。直观地说，失败能够提供信息：一个已经失败多次的问题更有可能是困难问题，因此再次尝试所带来的预期边际收益小于从新的实例开始。ReD 直接实现了这种“广度优先”的方法，并且可以轻松部署到任何已经支持重复采样和验证的流程中。

我们的工作证明了以下结论：(1) 我们通过更新理论将 pass@k 与 coverage@cost 联系起来，并推导出给定 pass@k 时 coverage@cost 关于预算的表达式。(2) 我们证明，对于任何预算和任何底层难度分布，重置都能严格提高 coverage@cost；最优的重置协议是每次尝试后都使用 ReD（$\tau = 1$）。(3) 给定经验或建模得到的 pass@k 曲线，我们可以预测 ReD 为达到任意目标覆盖率所节省的预算；反过来，当直接测量大规模-k pass@k 不切实际时，ReD 轨迹可提供一种统计效率高的方法来估计推理幂律指数。

我们在多个大语言模型和 Hu- manEval 基准测试（Chen 等人，2021）上验证了这些结果，证明 ReD 在固定预算下能够增加已解决的不同问题数量，相较于分配预算直至完成解题，可大幅节省成本。

我们的主要贡献如下：

- 我们推导出 pass@k 到 coverage@cost 的**精确映射**，并刻画其在幂律 pass@k 下的增长。具体而言，若 pass@k 按指数为 $0 < \alpha < 1$ 的幂律扩展（这在大语言模型中是常见的经验观察），标准的求解至完成方案会使 coverage@cost 呈次线性增长。然而，若 $\alpha > 1$，coverage@cost 则会线性增长。
- 我们提出**重置并丢弃（ReD）**，并证明**重置能在任何预算下提升 coverage@cost**，且**每次尝试都采用 ReD（$\tau = 1$）是最优的重置策略**。具体而言，若 pass@k 按指数为 $0 < \alpha < 1$ 的幂律扩展，ReD 会使 coverage@cost 线性增长；若 $\alpha > 1$，其增长仍为线性，且斜率高于求解至完成方案。
- 我们展示了如何根据 pass@k**预测尝试次数的节省**，以及如何在大规模 k 评估成本高昂时，根据 ReD 运行**估算推理扩展指数**。
- 我们通过实验证明，ReD 能显著提升可验证的大语言模型编码基准上的吞吐量。

本文其余部分安排如下：第3节，我们定义 coverage@cost，并给出它与 pass@k 指标之间的映射。第4节，我们针对所提出的重置策略（ReD），推导 coverage@cost 的主要理论结果，其中唯一任务的数量很大。我们还讨论了 ReD 在有限规模数据集上的实际实现。第5节，我们展示了如何利用 ReD 高效推断 pass@k 的标度指数。最后，我们在第6节对大语言模型编码任务中的结果进行实证验证，并在第7节作总结。

2. 相关工作

可验证任务上的测试时计算。 通过自动验证进行重复采样，是提升大语言模型在编程任务、数学和形式化证明搜索中性能的标准方法，通常用 pass@k (Chen 等人，2021）来概括。近期研究记录了在大规模采样预算下，覆盖率/pass@k 的推理时扩展情况（Brown 等人，2024年；Wu 等人，2024年），并探讨如何通过搜索和验证器在单个提示内以计算最优的方式分配测试时计算 （Snell 等人，2024年）；另见关于推理时算法的综述（Welleck 等人，2024年）。

推理时扩展定律与大规模预测k。 一些研究提出了用于解释或预测经验 pass@k 扩展的模型，包括简单的统计假设和难度混合解释（Levi，2025；Schaeffer 等人，2025）。由于大规模k 评估成本高昂，近期研究开发了高采样效率的估计器和自适应测量方案，以预测大规模k 行为（Kazdan 等人，2025）。与独立同分布重采样互补的是，动态分解通过自适应选择分解粒度来提升逐实例推理扩展（Light 等人，2025）。

预算感知评估与全局吞吐量。多项研究主张，应在明确的预算（查询数/词元数/美元）下评估推理策略，而非采用固定预算 k （Wang等人，2024年）。在合成数据和自我改进流程中，研究者也会追踪类似覆盖率的目标以及生成训练数据时的采样成本（Bansal等人，2025；Ding等人，2025）。在大语言模型领域之外，“覆盖率”（在时间/内存限制内解决的实例数量）是经典规划中的一项标准基准指标（Torralba等人，2020）。迭代代码修复/搜索中也存在相关的资源分配问题：在固定的模型调用预算下权衡广度与深度（Tang等人，2024年）。

重启、随机重置与重尾运行时间。重启策略常用于随机算法和组合搜索，以降低逐实例运行时间及其变异性（Luby 等人，1993；Gomes 等人，1997；2000；Walsh，1999；Williams 等人，2003；Streeter 等人，2007）。另一个研究方向则探讨了

随机过程和非平衡统计物理中的重启作用（Evans & Majumdar，2011；Gupta 等人，2014；Kusmierz 等人，2014；Belan，2018；Gupta 等人，2020；De Bruyne 等人，2020；2022；Sokolov，2023；Kumar & Pal，2023；Yin & Barkai，2023）。研究结果揭示了重启在何种情况下以及如何缩短平均首次到达（完成）时间（Reuveni，2016；Pal & Reuveni，2017；Chechkin & Sokolov，2018；Pal 等人，2019；2022；Evans 等人，2020）。近期进展大幅拓展了重置的应用，证明它能够加速高维扩散和稀有事件过程中的采样（Blumer 等人，2022；2024；Church 等人，2025；Keidar 等人，2025；Blumer & Hirshberg，2025）。在机器学习训练中，重启思想也体现在热重启学习率调度、标签噪声下基于重置的正则化，以及关于重置间隔的训练加速预测中（Loshchilov & Hutter，2016；Bae 等人，2025；Meir 等人，2025）。

大多数既有工作都在提升单实例成功率或对其进行测量。相反，我们优化的是全局吞吐量：如何在许多相互独立的问题之间分配固定的推理预算。我们的 ReD 策略将重启原则应用于跨任务场景，并在给定预算的覆盖率方面带来可证明的提升。

3. 将 pass@k 映射为 coverage@cost

3.1. pass@k 及其渐近幂律行为

我们采用与（Levi，2025；Schaeffer 等人，2025；Kazdan 等人，2025）类似的记号。在全文中，我们假设存在一个**完美验证器**；关于放宽这一假设的讨论见展望部分（第7节）。我们将对第 i个问题进行单次尝试时的成功概率记为 p_i，并将其累积分布函数（CDF）记为 $\text{pass}_i@k = 1-(1-p_i)^k$。每个问题都有不同的 p_i（“难度”），该值从某个分布 $\mathcal{P}(p)$ 中抽取。然后，对问题集合取平均后，pass@k 被定义为在 k 次试验中至少成功一次的概率，

$$\begin{aligned}\text{pass@}k &:= \mathbb{E}\left[\text{pass}_i@k\right] = \\ &= 1-\mathbb{E}\left[(1-p_i)^k\right] = 1-\int_0^1 (1-p)^k \mathcal{P}(p)\mathrm{d}p.\end{aligned} \tag{1}$$

公式（1）具有普遍性，在温和假设下会导出 pass@k 的渐近幂律行为。为此，我们考虑如下难度分布：对于 $p \ll 1$，其形式为$\mathcal{P}(p) \simeq cp^{\alpha-1}$，其中 c 和 α 均为正数。接着注意到 $1-\text{pass@}k = \frac{1}{k}\int_0^k (1-z/k)^k \mathcal{P}(z/k)\,\mathrm{d}z$，通过简单的变量替换 $z = pk$。取 k的极限可得

$$1-\text{pass@}k \underset{k\gg 1}{\approx} \frac{c}{k^\alpha}\int_0^\infty e^{-z} z^{\alpha-1}\mathrm{d}z = c\Gamma(\alpha)k^{-\alpha}, \tag{2}$$

其中我们使用了伽马函数的定义，$\Gamma(\cdot)$。由式（2）可以清楚地看出，任何“难度的分布”$\mathcal{P}(p)$在较小的 p处具有幂律行为，会导致 pass@k的幂律缩放，正如（Schaeffer 等人，2025）也得到的那样。

3.2. 根据 pass 预测 coverage@cost s@k

coverage@cost 指标表示在给定预算或尝试次数下已回答的不重复问题的平均数量。

定义 3.1。设 $x(t)$ 表示大语言模型在累计尝试次数达到t后已回答的不重复问题数量。每次实验中，问题池中的问题顺序都是随机的。模型会反复尝试解决每个问题，直至问题已解决。然后，

$$x(t) := \sum_{n=1}^{\infty} \mathbb{I}_{\{J_n \le t\}} = \sup\{n : J_n \le t\}, \tag{3}$$

其中，$\mathbb{I}$ 是指示函数，而 J_n 是回答问题 $i = 1, \ldots\ldots n$的总尝试次数，

$$J_n := \sum_{i=1}^{n} T^{(i)}, \tag{4}$$

其中 $T^{(i)}$ 是首次回答题库中第 i个问题的尝试次数。由于所有问题都是独立同分布的，$T^{(i)} \sim T,\ \forall i$。我们定义

$$\text{coverage@cost}(t) := \langle x(t)\rangle, \tag{5}$$

其中，我们保留记号 $\mathbb{E}[\cdot]$ 表示对问题难度分布取平均，并用括号 $\langle\cdot\rangle$ 表示对随机实现取平均。

我们认识到式（3）是一个更新过程，并且coverage@cost（t）可以用以下递推关系预测（Grimmett & Stirzaker，2020），

$$\begin{aligned}&\text{coverage@cost}(t) = F(t) + \\ &\sum_{j=1}^{t} \text{coverage@cost}(t-j)\left[F(j)-F(j-1)\right].\end{aligned} \tag{6}$$

在式（6）中，$F(t) := \Pr(T \le t) = \text{pass@}t$ 是该更新过程的累积分布函数（完整推导见附录A.1）。类似地，我们可以得到二阶矩 $x(t)$（见附录A.1）。

3.3. 大规模数据集的 coverage@cost 增长

回顾 Z 变换的定义，

$$\mathcal{Z}\{\text{coverage@cost}\}(z) = \sum_{t=0}^{\infty} z^t \text{coverage@cost}(t),$$

其中 $z \in [0,1)$。我们将其应用于公式（6），并在$z = 1$附近展开，以得到在以下时间点的 coverage@cost 行为

这些时间点大于数据集中回答问题的典型耗时。如附录 A.2所示，取决于推理幂律指数 α，可以观察到两种行为

$$\begin{cases}\text{coverage@cost}(t) \propto t^{\alpha} & 0<\alpha<1,\\ \text{coverage@cost}(t) \simeq \frac{t}{\mathbb{E}[T]} & \alpha>1,\end{cases} \tag{7}$$

也就是说，当 $0<\alpha<1, \mathbb{E}[T]$ 发散时，coverage@cost 将随 t 呈次线性增长。然而，如果 $\mathbb{E}[T]$ 有限，coverage@cost 将随 t 呈线性增长。

正如我们将在下一节中展示的，ReD 总能提高所有 t 的 coverage@cost。重要的是，对于 $\mathbb{E}[T]$ 发散的情况，ReD 会使其变为有限值，从而使 coverage@cost 发生显著的质变，从次线性增长转变为线性增长。对多种模型和数据集进行的分析（Levi, 2025）表明，更常见的情况是 $\alpha<1$，这意味着 $\mathbb{E}[T]$ 发散，凸显了 ReD 方法的优势。如果 $\mathbb{E}[T]$ 有限，ReD 将减少回答随机抽取问题所需的平均尝试次数，因此会提高 coverage@cost 随 t变化的线性增长斜率。

4. ReD 提升 coverage@cost

我们首先正式介绍算法 1 （重置与丢弃）。形式上，ReD 按轮次循环处理问题。我们对每个问题最多尝试 τ次。如果问题已解决，则将其加入覆盖集合并从问题池中丢弃；否则，我们通过将其移至队列末尾来进行重置，然后继续处理下一个问题。

算法 1 重置并丢弃（ReD）

输入： 问题池 $Q=\{q_i\}_{i=1}^{N}$，尝试极限 τ
$L \leftarrow$包含 Q 中所有问题的队列{initial order}
覆盖率 $\leftarrow \emptyset$ {solved questions}
当 L 不为空时 **执行**
  $q \leftarrow$ 从队列中取出(L) {pick the next question in line}
  已解决 $\leftarrow$ 否
  当 $T=1 \rightarrow \tau$ **执行**
    尝试问题 q
    如果 成功 **则**
      覆盖率 $\leftarrow$ 覆盖率 $\cup \{q\}$
      已解决 $\leftarrow$ True
      $Q \leftarrow Q \setminus \{q\}$ {永久丢弃 q}
      中断
    结束如果
  结束for循环
  如果未已解决 **则**
    入队(L, q) {reset: 将 q 发送到队尾}
  结束如果
结束while循环
返回 Q，覆盖率

4.1. 将 pass@k 与 coverage@cost 联系起来，使用 采用 ReD

公式（6）建立了 pass@k 指标与 coverage@cost 之间的联系。在这里，我们展示如何获得重置下的 coverage@cost，即当一个问题经过 τ 次尝试仍未得到回答时，将其替换为一个随机选择的新问题。我们首先考虑无限大的问题池，并在第 4.3节讨论有限规模数据集。该 ReD 流程由一个不同的随机变量 T_τ刻画，该变量表示在重置条件下正确回答一个随机选定问题所需的尝试次数，其定义如下：

$$T_\tau = \begin{cases} T & \text{if } T \leqslant \tau,\\ \tau + T'_\tau & \text{if } T > \tau,\end{cases} \tag{8}$$

其中，T'_τ是 T_τ的一个独立同分布副本，且与 T相互独立。

如果将随机变量 T 替换为 T_τ，式 (6) 也适用于 ReD 协议。式 (3描述了 ReD 过程。因此，只需将 $F(t)$ 替换为 $F_\tau(t)$，即 T_τ的累积分布函数。

$$\begin{aligned}&\text{coverage@cost}_\tau(t) = F_\tau(t) + \\ &\sum_{j=1}^{t} \text{coverage@cost}_\tau(t-j)\left[F_\tau(j) - F_\tau(j-1)\right].\end{aligned} \tag{9}$$

接下来，我们利用求解至完成策略的 $F(t) = \text{pass@}t$ 与 $F_\tau(t)$之间的已知联系（Eliazar &Reuveni, 2021）。

$$\begin{aligned}&F_\tau(t) = 1-(1-F(\tau))^n (1-F(u)),\\ &n = \left\lfloor \frac{t}{\tau} \right\rfloor, \quad u = t - n\tau.\end{aligned} \tag{10}$$

将式 (10) 代入式 (9)，即可预测$_\tau(t)$ 的 coverage@cost：每隔 τ时间间隔进行问题重置时，其值为 $F(k) = \text{pass@}k$ （不进行重置时）。

4.2. 重置总能提升 coverage@cost，且每次尝试都重置是最优策略

在间隔 τ、$\mathbb{E}[T_\tau]$进行重置时，随机选取的问题被正确回答所需的平均尝试次数，可通过计算以下求和得到：$\mathbb{E}[T_\tau] = \sum_{k=0}^{\infty}(1-F_\tau(k))$。代入式 (10)，可得（Eliazar& Reuveni, 2020; Bonomo & Pal, 2021; Meir 等人, 2025），

$$\mathbb{E}[T_\tau] = \frac{G(\tau)}{F(\tau)}, \quad G(\tau) = \sum_{k=0}^{\tau-1}(1-F(k)). \tag{11}$$

重要的是，我们得到，对于有限的 $\tau \geq 1$，$G(\tau)$ 和 $1/F(\tau)$ 都是有限的，因为对于任意 $t \geq 1$，$0 < F(t) \leq 1$。这意味着 $\mathbb{E}[T_\tau]$ 也是有限的。因此，只有最下面的

Eq. (7) 的那一行才相关。代入 Eq. (11) 后，coverage@cost 呈线性增长$_\tau(t)$，

$$\text{coverage@cost}_\tau(t) \simeq \frac{t}{\mathbb{E}[T_\tau]} = \frac{tF(\tau)}{G(\tau)}. \tag{12}$$

将该方程与公式（7）进行比较，可以立即看出，对于 $\alpha < 1$的情况使用 ReD 将大幅提升 coverage@cost。若不使用 ReD，则 coverage@cost 随时间呈次线性增长；而仅通过应用重置，它将呈线性增长，无论所使用的 τ 为何，或 $\alpha < 1$的具体取值为何。

对于 $\alpha > 1$的情况，此时求解至完成协议和 ReD 均呈线性增长，我们在附录B.1的推论B中证明，无论问题难度的底层分布 $\mathcal{P}$（p）如何，ReD 始终有益。

最后，我们在附录B 中证明，无论难度分布$\tau = 1$ 如何，最优重置时间均为 $\mathcal{P}$（p），具体总结如下定理。

定理4.1（ReD最优性）。 对于任意 $\mathcal{P}$（p）和 $\forall\tau \in \mathbb{N} \setminus \{0\}$, $\mathbb{E}[T_\tau] \leq \mathbb{E}[T_{\tau+1}]$, 即最优重置时间为 $\tau = 1$（证明见附录*B*）。

到目前为止，我们仅考虑了确定性的重置间隔，但一般而言，随机的重置时间可以从某个分布中采样。然而，已知确定性的 τ 重置可使重置条件下完成任务的平均时间最短（Pal & Reuveni，2017）。因此，定理4.1 也表明，每次尝试后都进行确定性重置是最优的重置策略。

4.3. 有限规模数据集的 coverage@cost 预测

为推导 ReD 下 coverage@cost 的方程，我们忽略了丢弃问题的影响，并假设问题池无限大，即重复回答同一个问题的概率为零。这里，我们推导由 N 个问题组成的有限规模数据集在重置间隔为 $\tau = 1$ 时，ReD 的 coverage@cost 表达式。在这种情形下，我们会经历多个轮次，并尝试各回答一次剩余问题。

我们首先利用 pass@k 求出经过 n 轮后剩余问题的数量 R_n。第 n 轮结束时未回答问题的平均数量为：

$$\langle R_n \rangle = N(1 - F(n)) = N(1 - \text{pass@}n). \tag{13}$$

因为一个问题在经过 n轮次后仍然存活的概率为 $1 - F(n)$。那么，

$$\text{coverage@cost}_{\tau=1}(t(n)) = N - \langle R_n \rangle = N\text{pass@}n, \tag{14}$$

其中 $t(n)$是第n轮次结束时的总尝试次数。进而，总尝试次数由下式给出，

$$t(n) = \sum_{k=0}^{n-1} R_k. \tag{15}$$

取平均值，并将式（13）代入式（15），得到

$$\langle t(n) \rangle = N \sum_{k=0}^{n-1} (1 - F(k)) = N \sum_{k=0}^{n-1} (1 - \text{pass@}k). \tag{16}$$

为预测有限数据集上的 coverage@cost$_{\tau=1}$，我们将式（$t(n)$）中的14替换为其平均值 $\langle t(n) \rangle$，从而得到以下近似

$$\text{coverage@cost}_{\tau=1}(\langle t(n) \rangle) \approx N\text{pass@}n, \tag{17}$$

用于固定时间点 $t_n = \langle t(n) \rangle$处的 coverage@cost。该近似忽略了 R_n 和 t_n中的涨落。我们在第6 节展示了该近似对大语言模型效果良好。

5. 使用 ReD 根据 coverage@cost 推断 pass@k 幂律指数

5.1. ReD 改变难度分布

在第4节中，我们构建了一个框架，用于在已知 pass@k 的情况下预测 coverage@cost。这里，我们关注 pass@k 未知的情形，并展示 ReD 策略能够在不完整评估 pass@k 的情况下推断 pass@k的缩放律指数。

我们首先考察问题难度的分布$\mathcal{P}(p)$。在 ReD 协议中，由于我们会丢弃之前轮次中已答对的问题，因此每一轮的问题难度分布实际上都会发生变化。我们将第 n 轮的分布记为 $\mathcal{P}_n(p)$。下面将证明，只要对于 $p \ll 1$, 初始分布 $\mathcal{P}_0(p) \simeq cp^{\alpha-1}$，经过若干轮后，$\mathcal{P}_n(p)$ 就必然趋近于贝塔分布。随后，我们假设 $\mathcal{P}_0(p) \sim \text{Beta}(p;$ α, $\beta)$，忽略初始松弛阶段，并据此推导 coverage@cost$_{\tau=1}(t)$ 的结果。

5.2. 难度渐近收敛至贝塔分布

分布 $\mathcal{P}_n(p)$ 表示经过 n 轮后仍未得到回答的所有问题。这由下式给出

$$\mathcal{P}_n(p) = \frac{\mathcal{P}_0(p)(1-p)^n}{\int_0^1 (1-p)^n \mathcal{P}_0(p) \mathrm{d}p}, \tag{18}$$

其中，难度为 p的问题的新权重由其初始权重 $\mathcal{P}_0(p)$ 与存活 n轮次的概率（$1-p$）^{n}相乘得到。请注意，当 n取值较大时，项（$1-p$）^{n} 会抑制较高取值的贡献

p。在此极限下，可以安全地近似为 $\mathcal{P}_0(p) \simeq cp^{\alpha-1}$，从而得到

$$\mathcal{P}_n(p) \underset{n\gg 1}{\approx} \frac{p^{\alpha-1}(1-p)^n}{\int_0^1 p^{\alpha-1}(1-p)^n \mathrm{d}p} \sim \mathrm{Beta}(p;\alpha,n+1). \tag{19}$$

5.3. 从 ReD 推断幂律指数

以往研究（Brown 等人，2024年；Schaeffer 等人，2025；Kazdan 等人，2025）提出了多种方法来获得 α 的 1 –pass@k 标度指数。本文推导了一种适用于 ReD 过程的估计方法。我们首先求得适用于 $\mathcal{P}_n(p)$ 所有取值的 n。

忽略向贝塔分布的弛豫，并使用5.2 节中的 $\mathcal{P}_0(p) = \mathrm{Beta}(p;\ \alpha,\ \beta)$ 代入式（18），重复推导过程，可得：

$$\mathcal{P}_n(p) = \mathrm{Beta}(p;\alpha,\beta+n). \tag{20}$$

式（20）的含义是，平均而言，$\gamma_n R_n$ 个问题将在第 n 轮得到正确解答，其中$\gamma_n :=\alpha/(\alpha+\beta+n)$ 是 $\mathrm{Beta}(p;\ \alpha,\ \beta+n)$的均值。因此，该过程的动态可写为：

$$\langle R_{n+1}\rangle = \langle R_n\rangle - \gamma_n\langle R_n\rangle. \tag{21}$$

通过重新整理公式（21），并利用 γ_n的定义，我们得到一个关于轮次数的线性方程，

$$-\frac{\langle R_n\rangle}{\langle R_{n+1}-R_n\rangle} = \frac{1}{\alpha}n + \frac{\alpha+\beta}{\alpha}. \tag{22}$$

因此，将比值 $\langle R_n\rangle/\langle R_{n+1}-R_n\rangle$ 对 n 作图，可得一条斜率为 $1/\alpha$ 的直线。经过几轮后，便可通过线性拟合提取标度指数 α。

6. 大语言模型实验

为展示 ReD 在大语言模型上的效率，我们在 HumanEval 数据集（Chen 等人，2021）上进行了实验。我们采用了两种查询协议：第一种是解题直至完成的方法，即反复提问，直到问题得到正确回答后再转向下一个问题。第二种是 ReD 方法，每次尝试后都重置，相当于分轮次提问：每轮将所有问题都问一遍，并在下一轮中舍弃已正确回答的问题。

在实践中，我们通过对每个模型评估 pass@k ，最多评估 $k = 100$ 次，来进行这些实验；并针对每个问题和每次尝试，将是否答对记录在一个维度为“问题数×尝试次数”的结果矩阵中。我们还记录了使用了多少输入和输出词元。为了生成许多求解至完成评估过程的随机实现，我们对结果矩阵的行和列进行打乱，并将其视为采用了求解至完成协议来分析。我们还基于同一份打乱后的数据生成 ReD 协议的所有实现：按轮次分析结果矩阵，逐列处理，并丢弃此前轮次中已答对问题所对应的行。

我们使用 Groq API 和三种不同的模型进行实验：llama-3.1-8b-instant、llama-3.3-70b-versatile 和 openai/gpt-oss-20b（OpenAI，2025）。三个模型使用了相同的提示。我们设置了严格的系统提示词，以减少对话中的冗余内容，并确保输出格式一致：“你是一名智能编程助手。系统将提供一个函数签名和文档字符串。你必须补全函数体。仅返回该函数实现的 *Python* 代码。不要添加解释或测试。请将代码放在 *Python Markdown* 代码块中。”随后，我们使用了标准 HumanEval 提示词，其中包含函数签名和文档字符串。我们使用正则表达式提取输出代码。所有实验的温度均设为 0.8。本文用于生成结果的所有代码都将在 GitHub 上发布。

图 1 展示了三个模型分别使用求解至完成协议和 ReD 时，coverage@cost（按问题总数归一化）随累计尝试次数变化的比较结果。对于每个模型，我们绘制了 100 次随机实现的平均覆盖率（虚线）及其标准差（阴影区域）。我们还展示了每种评估协议所使用的随机实现。对于所有模型，平均而言，使用 ReD 时的 coverage@cost 始终高于求解至完成协议，同时能够达到相同的渐近覆盖率。

另一个值得探讨的问题是：给定一个数据集，哪种策略更好？是使用结合 ReD 的小模型，还是训练一个大得多的模型？我们在图2中回答了这一问题：比较较小的 llama-3.1-8b-instant 模型分别使用求解至完成协议和 ReD 时的 coverage@cost，并将其与较大模型使用求解至完成协议时的结果进行比较。我们采用三种不同的成本衡量方式：回答问题所需的尝试总次数、使用的累计词元数，以及以美元计的预算。词元到美元的换算依据表 1中的价格进行。要全面了解 ReD 相较于标准方法的效率，考察这三种成本衡量方式非常重要。

从尝试总次数来看，在覆盖率达到 $\sim$ 90%之前，对较小模型使用 ReD 比较大的 Llama模型更高效；而在覆盖率达到 $\sim$ 80%之前，ReD 的表现优于或不逊于 gpt-oss-20b 模型。显然，两个较大的模型最终也能正确回答小模型无法回答的问题，

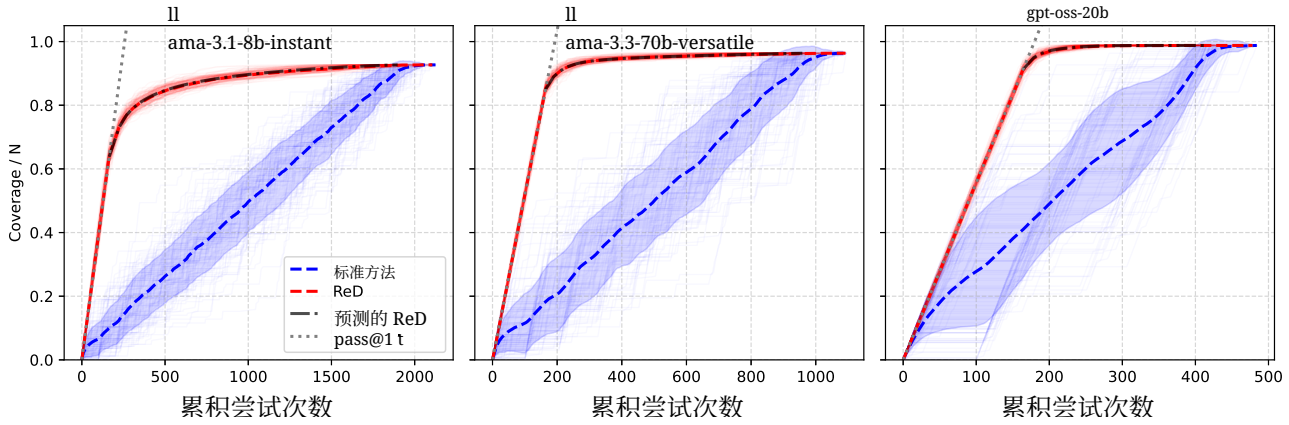


图 *1.* coverage@cost（按问题总数归一化），$N = 164$，展示了三个模型在 ReD 和标准方法（求解至完成）评估协议下，coverage@cost 随累积尝试次数的变化情况：（左）llama-3.1-8b-instant、（中）llama-3.3-70b-versatile 和（右）gpt-oss-20b。虚线和阴影区域分别表示随机实现的均值和标准差。点划线表示使用公式（16）和（17）得出的预测。点线表示 pass@$1 \cdot t$，这是第一轮 ReD 中的预期表现。

但这需要投入大量训练工作。ReD 可以利用小模型根据问题难度对问题进行有效排序，将最难的问题留给较大的模型。

从词元数量来看，模型的输出冗长程度起着重要作用。在这种情况下，对于几乎所有准确率水平，较小的 Llama 模型采用 ReD 的表现都优于 gpt-oss-20b。原因在于 gpt-oss-20b 模型的输出非常冗长，因而会使用大量词元。通过提示工程尽可能降低这部分额外词元成本似乎是可行的，但为了进行公平比较，我们对所有模型使用了相同的提示。较大的 Llama 模型输出远不如 gpt-oss-20b 冗长，因此在覆盖率与词元数量的关系上表现更好。不过，即便如此，在覆盖率达到 $\sim 80\%$ 之前，ReD 仍是更优策略。

以美元计，ReD 在所有情况下显然都是更好的策略，因为较大的模型要么非常消耗词元，要么非常昂贵（见表 1）。首选策略是先使用 ReD 回答所有能够正确回答的问题，然后仅针对它完全未能回答的问题切换到较大的模型。

模型	**输入价格** （每 100 万词元）	**输出价格** （每 100 万词元）
llama-3.1-8b-instant	0.05 美元	0.08 美元
llama-3.3-70b-versatile	0.59 美元	0.79 美元
gpt-oss-20b	$0.075	$0.30

表1. 用于计算美元成本的 Groq 词元定价。

最后，我们在图 3 中展示了如何使用 ReD 推断较小 Llama模型的幂律指数。

我们证明，使用公式 (22，可以通过对 $\langle R_n \rangle / \langle R_{n+1} - R_n \rangle$与 n进行线性拟合，正确推断出幂律行为。得到的幂律为 0.34 ± 0.01，与根据 k 在 k 较高时的表现估算出的幂律相符，其中 0。34 （见图 S1）。请注意，只要 $R_n \gg 1$，线性回归就能保持准确；否则， $R_{n+1} - R_n$ 的倒数会引入较大噪声。因此，线性拟合只进行到第 15 轮。本文未展示两个较大模型的幂律指数推断结果，因为它们使用 ReD 时只需很少几轮就能解答几乎所有问题。

7. 总结与展望

我们提出了重置并丢弃（ReD），这是一种针对大语言模型的新型重置式查询协议，可在固定预算下提升 coverage@cost。我们将 pass@k 映射为 coverage @cost，并证明，pass@k 在 $0 < \alpha < 1$下呈现幂律行为，会使 coverage@cost 呈次线性增长。我们证明 ReD 始终能够提升 coverage@cost；尤其是在 $0 < \alpha < 1$的情况下，ReD 可使 coverage@cost 线性增长，而最优重置协议是在每次尝试时都采用 ReD。我们还表明，给定 pass@k，即可预测每次尝试都执行 ReD 所带来的节省金额。如果无法获得 pass@k，我们证明可以利用 ReD 估算幂律指数，而无需进行完整的 pass@k 评估。最后，我们展示了 ReD 在 HumanEval 编程任务数据集上的表现，测试对象为三个不同规模的大语言模型。在按尝试次数、累计令牌数或美元计量的广泛成本范围内，ReD 的表现均优于标准求解至完成基线。

ReD 的实现对大语言模型的商业部署有着深远影响，具体而言，它能够促成从按词元付费转向“按结果付费”——

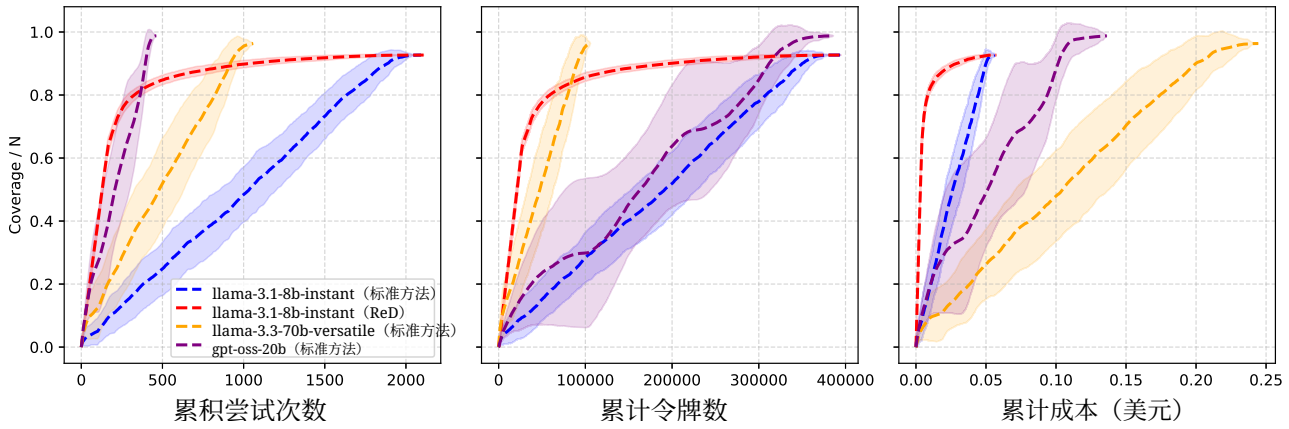


图 *2.* llama-3.1-8b-instant（标准方法、求解至完成与 ReD）的 coverage@cost 按问题总数 $N = 164$ 归一化，并与 llama-70b-versatile 和 gpt-oss-20b（仅采用标准方法、求解至完成）进行比较；测量指标为（左）尝试次数、（中）词元和（右）美元成本。虚线和阴影区域分别表示随机实现的均值和标准差。

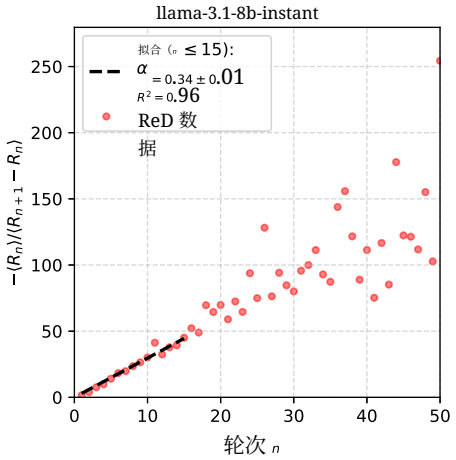


图 *3.* llama-3.1-8b-instant 的比率 $-\langle R_n \rangle / \langle R_{n+1} - R_n \rangle$ 随轮次 n 的变化，其中 R_n 是第 n 轮时未解决问题的数量。

。目前，商业 API 的用户承担着模型失败的经济风险，无论生成结果是否有效，都必须为生成付费。通过将 ReD 用作后端分配策略，类似于 FrugalGPT（Chen 等人，2024b）等分层路由方案，服务提供商可以转而销售已验证解答。这样便能协调各方激励：用户只需为已解决的任务支付溢价，而服务提供商则通过优化“有效吞吐”（每美元解决的有效任务数）来实现利润最大化；ReD 能在不让用户接触底层搜索复杂性的情况下，有效权衡模型成本与任务难度。

除了商业服务之外，ReD 还显著提升了研究和边缘计算领域的效率。在 STaR 等强化学习算法中（Zelikman 等人，2022）或 ReST（Gulcehre 等人，2023年），这些算法依赖自生成合成数据，ReD 可以充当主动学习筛选器：与其将 GPU 小时均匀浪费在简单或不可能解决的问题上，不如让 ReD 将计算预算集中在“可学习性的前沿”，以相同的训练成本获得更多成功轨迹。最后，ReD 可用于决定先将哪些问题交给较小的模型处理，以及将哪些问题留给更大、成本更高的模型。从这个意义上说，它能自动优化智能水平与成本之间的权衡。这对于移动与边缘应用尤为重要（Bansal 等人，2025），在这些场景中，ReD 为资源受限推理提供了一个形式化框架，准确判定何时应停止尝试处理某个查询，例如在最大化为用户带来的总效用的同时，保留电池续航。

局限性与未来工作： 在这项工作中，我们仅考虑了完全可验证任务。验证仍是一个开放问题，值得研究在不同验证方法下的 ReD，例如 Best-of-N (Wang 等人，2023年）、多数投票和奖励模型（Christiano 等人，2023年）。

此外，我们将独立试验的并行采样方案作为充分的基线。然而，研究表明，在许多情况下，进行序贯采样（或回溯）更有益（Ku-mar et al.，2024年；Qin et al.，2025；Muennighoff etal.，2023年）。未来工作中，一个自然的方向是将 ReD 框架扩展到相关推理尝试，使先前状态的信息能够在整个推理轨迹中得以保留。

致谢

B.H. 感谢以色列科学基金会（资助编号：1037/22 和 1312/22）以及 Pazy 基金会

IAEC-UPBC（资助编号：415-2023）提供的支持。本项目获得欧盟地平线2020研究与创新计划下欧洲研究理事会（ERC）的资助（授予 S.R. 的资助协议编号：947731）。S.M. 感谢特拉维夫大学人工智能与数据科学中心（TAD）以及以色列规划与预算委员会（VATAT）的支持。N.L. 获得 EPFL AI4science/AI中心计划的支持。

参考文献

Bae, Y., Song, Y., and Jeong, H. 随机重置缓解了标签噪声导致的随机梯度下降潜在梯度偏差。《机器学习：科学与技术》，6(1):015062, 2025年。网址：https://doi.org/10.1088/2632-2153/adbc46。Bansal, H., Hosseini, A., Agarwal, R., Tran, V. Q., and Kazemi, M. 更小、更弱，却更出色：通过计算最优采样训练大语言模型推理器。载于第十三届国际学习表征会议，2025年。网址：https://openreview.net/forum?id=3OyaXFQuDl。Belan, S. 重启或可优化伯努利试验的成功概率。《物理评论快报》, 120(8): 080601, 2018年。网址：https://doi.org/10.1103/PhysRevLett.120.080601。Blumer, O. and Hirshberg, B. 你试过关掉再重启吗？用于增强采样的随机重置。《威利跨学科评论：计算分子科学》，15(4):e70038, 2025年。网址：https://doi.org/10.1002/wcms.70038。Blumer, O., Reuveni, S., and Hirshberg, B. 用于增强采样的随机重置。《物理化学快报》, 13(48):11230–11236, 2022。网址：https://pubs.acs.org/doi/10.1021/acs.jpclett.2c03055。Blumer, O., Reuveni, S., and Hirshberg, B. 用于改进动力学推断的短时低频元动力学。《化学理论与计算杂志》, 20(9):3484–3491, 2024年。网址：https://pubs.acs.org/doi/10.1021/acs.jctc.4c00170。Bonomo, O. L. and Pal, A. 离散空间与时间下重启时的首次通过：在一维受限格点随机游走中的应用。《物理评论*E*》，103:052129, 2021年5月。doi: 10.1103/PhysRevE.103.052129。网址：https://doi.org/10.1103/PhysRevE.103.052129。Brown, B., Juravsky, J., Ehrlich, R. S., Clark, R., Le, Q. V., Ré, C., and Mirhoseini, A. 大语言模型作者：通过重复采样扩展推理计算。*arXiv*预印本 *arXiv:2407.21787*，2024年。网址：https://arxiv.org/abs/2407.21787。Chechkin, A. 和 Sokolov, I. M. 带重置的随机搜索：一种统一更新方法。《物理评论快报》，121(5):050601，2018年。网址：https://doi.org/10.1103/PhysRevLett.121.050601。Chen, L., Davis, J. Q., Hanin, B., Bailis, P., Stoica, I., Zaharia, M. 和 Zou, J. 更多大语言模型调用是否就是全部所需？迈向复合推理系统的扩展定律，2024a。网址 https://arxiv.org/abs/2403.02419。Chen, L., Zaharia, M. 和 Zou, J. FrugalGPT：如何在降低成本的同时使用大语言模型并提升性能。《机器学习研究汇刊》，2024b。ISSN 2835-8856。网址 https://openreview.net/forum?id=cSimKw5p6R。特色认证。Chen, M., Tworek, J., Jun, H., Yuan, Q., de Oliveira Pinto, H. P., Kaplan, J., Edwards, H., Burda, Y., Joseph, N., Brockman, G., Ray, A., Puri, R., Krueger, G., Petrov, M., Khlaaf, H., Sastry, G., Mishkin, P., Chan, B., Gray, S., Ryder, N., Pavlov, M., Power, A., Kaiser, L., Bavarian, M., Winter, C., Tillet, P., Such, F. P., Cummings, D., Plappert, M., Chantzis, F., Barnes, E., Herbert-Voss, A., Guss, W. H., Nichol, A., Paino, A., Tezak, N., Tang, J., Babuschkin, I., Balaji, S., Jain, S., Saunders, W., Hesse, C., Carr, A. N., Leike, J., Achiam, J., Misra, V., Morikawa, E., Radford, A., Knight, M., Brundage, M., Murati, M., Mayer, K., Welinder, P., McGrew, B., Amodei, D., McCandlish, S., Sutskever, I. 和 Zaremba, W. 评估基于代码训练的大语言模型，2021年。网址 https://arxiv.org/abs/2107.03374。Christiano, P., Leike, J., Brown, T. B., Martic, M., Legg, S. 和 Amodei, D. 基于人类偏好的深度强化学习，2023年。网址 https://arxiv.org/abs/1706.03741。Church, J. R., Blumer, O., Keidar, T. D., Ploutno, L., Reuveni, S. 和 Hirshberg, B. 通过有信息重置加速分子动力学。《化学理论与计算杂志》，21(2):605–613，2025。网址 https://pubs.acs.org/doi/10.1021/acs.jctc.4c01238。De Bruyne, B., Randon-Furling, J. 和 Redner, S. 首达重置中的优化。《物理评论快报》，125(5):050602，2020。网址 https://doi.org/10.1103/PhysRevLett.125.050602。

De Bruyne, B., Majumdar, S. N. 和 Schehr, G. 通过增强涨落实现布朗桥的最优重置。《物理评论快报》，128(20): 200603，2022年。网址 https://doi.org/10.1103/PhysRevLett.128.200603。

Ding, Y., Xi, Z., He, W., Li, Z., Zhai, Y., Shi, X., Cai, X., Gui, T., Zhang, Q. 和 Huang, X. 通过苏格拉底式引导采样缓解大语言模型自我改进中的尾部变窄。在*2025*年美洲计算语言学协会分会年会论文集（*NAACL*），2025年。网址：https://aclanthology.org/2025.naacl-long.533/。arXiv:2411.00750。Ehrlich, R., Brown, B., Juravsky, J., Clark, R., Ré, C. 和 Mirhoseini, A. Codemonkeys：扩展软件工程的测试时计算，2025年。网址：https://arxiv.org/abs/2501.14723。Eliazar, I. 和 Reuveni, S. 尖锐重启的平均性能 I：统计路线图。《物理学杂志*A*：数学与理论》，53(40):405004，2020年。网址：https://doi.org/10.1088/1751-8121/abae8c。Eliazar, I. 和 Reuveni, S. 尖锐重启的尾部行为路线图。《物理学杂志*A*：数学与理论》，54(12):125001，2021年。网址：https://doi.org/10.1088/1751-8121/abe4a0。Evans, M. R. 和 Majumdar, S. N. 随机重置下的扩散。《物理评论快报》，106:160601，2011年。doi: 10.1103/PhysRevLett. 106.160601。网址：https://doi.org/10.1103/PhysRevLett.106.160601。Evans, M. R.、Majumdar, S. N. 和 Schehr, G. 随机重置及其应用。《物理学杂志*A*：数学与理论》，53(19):193001，2020年。网址：https://doi.org/10.1088/1751-8121/ab7cfe。Feller, W.《概率论及其应用导论》，第二卷，第2卷。John Wiley & Sons，1991年。First, E., Rabe, M. N., Ringer, T. 和 Brun, Y. Baldur：利用大语言模型生成和修复完整证明，2023年。网址：https://arxiv.org/abs/2303.04910。Gomes, C. P.、Selman, B. 和 Crato, N. 组合搜索中的重尾分布。载于约束规划的原理与实践（*CP*），1997年。网址：https://doi.org/10.1007/BFb0017434。

Gomes, C. P., Selman, B. 和 Kautz, H. A. 可满足性与约束满足问题中的重尾现象。《自动推理杂志》，24:67– 100，2000年。网址 https://doi.org/10.1023/A:1006314320276。Grimmett, G. 和 Stirzaker, D.概率与随机过程。牛津大学出版社，2020年。Gulcehre, C., Paine, T. L., Srinivasan, S., Kapturowski, Steven andGV, F., Denil, M., Hoffman, M. 和 Pascanu, R. 面向语言建模的强化自训练（ReST）。*arXiv*预印本 *arXiv:2308.08998*，2023年。网址 https://arxiv.org/abs/2308.08998。Gupta, D., Plata, C. A. 和 Pal, A. 随机重置中的功涨落与Jarzynski等式。《物理评论快报》，124(11):110608，2020年。网址 https://doi.org/10.1103/PhysRevLett.124.110608。Gupta, S., Majumdar, S. N. 和 Schehr, G. 受随机重置影响的涨落界面。《物理评论快报》，112(22):220601，2014年。网址 https://doi.org/10.1103/PhysRevLett.112.220601。Hassid, M., Remez, T., Gehring, J., Schwartz, R. 和 Adi, Y. 越大越好吗？通过预算重新分配改进大语言模型代码生成。发表于第一届语言建模会议，2024年。网址 https://openreview.net/forum?id=QJvfpWSpWm。Hughes, J., Price, S., Lynch, A., Schaeffer, R., Barez, F., Koyejo, S., Sleight, H., Jones, E., Perez, E. 和 Sharma, M. Best-of-n越狱，2024年。网址 https://arxiv.org/abs/2412.03556。Kazdan, J., Schaeffer, R., Allouah, Y., Sullivan, C., Yu, K., Levi, N. 和 Koyejo, S. 大语言模型中pass@k扩展的高效预测。*arXiv*预印本 *arXiv:2510.05197*，2025年。网址 https://arxiv.org/abs/2510.05197。Keidar, T. D., Blumer, O., Hirshberg, B. 和 Reuveni, S. 用于知情搜索策略的自适应重置及非平衡稳态的设计。《自然通讯》，16(1):7259，2025年。网址 https://doi.org/10.1038/s41467-025-62398-2。Kulal, S., Pasupat, P., Chandra, K., Lee, M., Padon, O., Aiken, A. 和 Liang, P. S. Spoc：基于搜索的伪代码转代码。见 Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E. 和 Garnett, R.（编），神经信息处理系统进展，第32卷。Curran Associates, Inc.，2019年。网址 https://proceedings.neurips.cc/paper_files/paper/2019/file/7298332f04ac004a0ca44cc69ecf6f6b-Paper.pdf。

Kumar, A. 和 Pal, A. 重启条件下记录年龄的通用框架。《物理评论快报》，130(15): 157101，2023年。网址 https://doi.org/10.1103/PhysRevLett.130.157101。

Kumar, A., Zhuang, V., Agarwal, R., Su, Y., Co-Reyes, J. D., Singh, A., Baumli, K., Iqbal, S., Bishop, C., Roelofs, R., Zhang, L. M., McKinney, K., Shrivas- tava, D., Paduraru, C., Tucker, G., Precup,D., Behba- hani, F., and Faust, A. 通过强化学习训练语言模型进行自我纠正，2024年。URLhttps://arxiv.org/abs/2409.12917。

Kusmierz, L., Majumdar, S.N., Sabhapandit, S., andSchehr, G. 莱维飞行重置下最优搜索时间的一级相变。*Phys. Rev. Lett.*，113: 220602，2014年11月。doi: 10.1103/PhysRevLett.113. 220602。
U
R
L
https://link.aps.org/doi/10.1103/PhysRevLett.113.220602。

Kwok, J., Agia, C., Sinha, R., Foutter, M., Li, S., Stoica, I., Mirhoseini, A., and Pavone, M. RoboMonkey：扩展视觉语言-动作模型的测试时采样与验证，2025 年。网址 https://arxiv.org/abs/2506.17811。

Levi, N. I. 推理缩放定律的一个简单模型。载于 第四十二届国际机器学习会议，2025 年。网址 https://openreview.net/forum?id=5ulCAfxJzc。

Light, J., Cheng, W., Yue, W., Oyamada, M., Wang, M., Paternain, S., and Chen, H. DISC：动态分解改进大语言模型推理缩放。*arXiv* 预印本 *arXiv:2502.16706*，2025 年。网址 https://arxiv.org/abs/2502.16706。

Loshchilov, I. 和 Hutter, F. SGDR：采用热重启的随机梯度下降。*arXiv* 预印本 *arXiv:1608.03983*，2016。网址https://arxiv.org/abs/1608.03983。

Luby, M.、Sinclair, A. 和 Zuckerman, D. 拉斯维加斯算法的最优加速。《信息处理快报》，47(4):173–180，1993。网址 https://doi.org/10.1016/0020-0190(93)90029-9。

Meir, S.、Keidar, T. D.、Reuveni, S. 和 Hirshberg, B. 用于优化扰动以改善机器学习模型训练的首次通过方法。《机器学习：科学与技术》，6(2):025053，2025。网址 https://doi.org/10.1088/2632-2153/add8df。

Metzler, R. 和 Klafter, J. 《随机游走：反常扩散指南：一种分数阶动力学方法》。《物理学报告》，339(1):1–77，2000年。

Muennighoff, N.、Rush, A.、Barak, B.、Le Scao, T.、Tazi, N.、Piktus, A.、Pyysalo, S.、Wolf, T. 和 Raffel, C. A. 受数据约束的语言模型的规模扩展。神经信息处理系统进展，36:50358–50376，2023年。OpenAI。gpt-oss-120b 和 gpt-oss-20b 模型卡，2025。网址：https://arxiv.org/abs/2508.10925。Pal, A. 和 Reuveni, S. 重启下的首次通过。《物理评论快报》，118:030603，2017年。doi：10.1103/PhysRevLett.118.030603。

Pal, A., Eliazar, I., 和 Reuveni, S. 带分支的重启下首次通过。《物理评论快报》，122(2): 020602, 2019。网址 https://doi.org/10.1103/PhysRevLett.122.020602。

Pal, A., Kostinski, S., 和 Reuveni, S. 随机重置中的检查悖论。《物理学杂志*A*：数学与理论》,55(2):021001,2022。

Qin, T., Alvarez-Melis, D., Jelassi, S., 和 Malach, E. 是否回溯：序贯搜索何时会限制模型推理，2025。网址 https://arxiv.org/abs/2504.07052。

Reuveni, S. 最优随机重启使首次通过时间的涨落具有普适性。《物理评论快报》，116(17):170601, 2016。doi: 10.1103/PhysRevLett. 116.170601。网址 https://doi.org/10.1103/PhysRevLett.116.170601。

Schaeffer, R., Kazdan, J., Hughes, J., Juravsky, J., Price, S., Lynch, A., Jones, E., Kirk, R., Mirhoseini, A., 和 Koyejo, S. 大型语言猴子如何获得它们的幂律？*arXiv* 预印本 *arXiv:2502.17578*，2025。

Scher, H. 和 Montroll, E. W. 非晶固体中的反常渡越时间弥散。物理评论 *B*，12:2455–2477，1975年9月。doi: 10.1103/PhysRevB. 12.2455。网址 https://doi.org/10.1103/PhysRevB.12.2455。

Snell, C., Lee, J., Xu, K., 和 Kumar, A. 对大语言模型的测试时计算进行最优扩展，可能比扩展模型参数更有效。*arXiv* 预印本 *arXiv:2408.03314*，2024年。网址https://arxiv.org/abs/2408.03314。

Sokolov, I. M. 重置条件下布朗运动的线性响应与涨落耗散关系。物理评论快报，130(6):067101，2023年。网址 https://doi.org/10.1103/PhysRevLett.130.067101。

Streeter, M., Golovin, D., and Smith, S. 问题实例集合的重启计划。载于*AAAI*人工智能大会论文集（*AAAI*），2007年。

Tang等人. 使用大语言模型进行代码修复会带来探索与利用之间的权衡。载于神经信息处理系统进展，2024年。Torralba, Á., Seipp, J., and Sievers, S. 面向经典规划的基准测试集自动配置。载于*ICAPS 2020*领域无关规划启发式与搜索研讨会（*HSDIP*），第58–66页，2020年。网址：
h
t
t
p
s://
a
i.
d
m
i.
u
n
i
b
a
s.
ch/papers/torralba-et-al-icaps2020wshsdip.pdf。Walsh, T. 小世界中的搜索。载于第十六届国际人工智能联合会议（*IJCAI*）论文集，1999年。Wang, J., Jain, S., Zhang, D., Ray, B., Kumar, V., and Athiwaratkun, B. 令牌经济中的推理：大语言模型推理策略的预算感知评估。载于*2024*年自然语言处理实证方法会议（*EMNLP*）论文集，2024年。网址：
h
ttps://aclanthology.org/2024.emnlp-main.1112/。Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A., and Zhou, D. 自洽性提升了语言模型中的思维链推理，2023年。网址：
https://arxiv.org/abs/2203.11171。Welleck, S., Bertsch, A., Finlayson, M., Schoelkopf, H., Xie, A., Neubig, G., Kulikov, I., and Harchaoui, Z. 从解码到元生成：大语言模型的推理时算法。《机器学习研究汇刊》，2024年10月。网址：
https://arxiv.org/abs/2406.16838。Williams, R., Gomes, C. P., and Selman, B. 论重尾性、后门与重启之间的联系。载于可满足性测试理论与应用（*SAT*），2003年。Wu, Y., Sun, Z., Li, S., Welleck, S., and Yang, Y. 推理扩展定律：对语言模型问题求解的计算最优推理进行实证分析，2024年。网址：https://arxiv.org/abs/2408.00724。Yin, R. a
nd Barkai, E. 重启可缩短量子游走的命中时间。《物理评论快报》，130(5)：050802，2023年。网址：
h
ttps://doi.org/10.1103/PhysRevLett.130.050802。Zelikman, E., Wu, Y., Mu, J., and Goodman, N. D. STaR：以推理自举推理。载于神经信息处理系统进展，第35卷，第15476–15488页，2022年。

附录

A. 将大规模数据集上的 coverage@cost 与 pass@k 联系起来

在本附录中，我们回顾更新理论中的已知结果（Grimmett & Stirzaker，2020）。为求完整并便于阅读，我们在下文重述相关证明。

A.1. 式（6）的证明

在正文中，我们将 $x(t)$ 定义如下：

$$x(t) \coloneqq \sum_{n=1}^{\infty} \mathbb{I}_{\{J_n \le t\}} = \sup\{n : J_n \le t\}, \quad J_n \coloneqq \sum_{i=1}^{n} T^{(i)}, \tag{S1}$$

其中，$T^{(i)}$ 是服从 T 分布的独立同分布正随机变量。已知第一个问题在第 j 步得到正确回答，更新过程满足以下递推关系：

$$x(t) \overset{d}{=} \begin{cases} 1 + x(t-j) & \text{if } j \le t, \\ 0 & \text{if } j > t. \end{cases} \tag{S2}$$

换言之，对于 $t < j$，我们有 $x(t) = 0$；否则，$x(t) \overset{d}{=} 1 + x(t-j)$。这是因为第一次成功发生在时间 j，此后过程重新开始，从这一时刻起的额外成功次数服从 $x(t-j)$。

为证明式（6），我们引入以下记号：$m_1(t) \coloneqq \text{coverage@cost}(t)$ 以及 $f(j) \coloneqq F(j) - F(j-1)$，其中$F(j)$ 是 T的累积分布函数，且 $f(j)$ 是其概率质量函数。注意，采用此记号后，式（6）可写为

$$m_1(t) = F(t) + \sum_{j=1}^{t} m_1(t-j) f(j). \tag{S3}$$

接下来，我们结合全期望定律，使用式（S2）中的递推关系，$\langle x \rangle = \langle \langle x \mid y \rangle \rangle$，其中 $\langle x \mid y \rangle$ 是 x 在给定 y条件下的期望。

$$\begin{aligned} m_1(t) &= \langle x(t) \rangle = \langle \langle x(t) \mid T^{(1)} \rangle \rangle = \sum_{j=1}^{\infty} \langle [x(t) \mid T^{(1)} = j] \rangle f(j) = \sum_{j=1}^{\infty} \mathbb{I}_{\{j \le t\}} (1 + \langle x(t-j) \rangle) f(j) \\ &= \sum_{j=1}^{t} (1 + m_1(t-j)) f(j) = F(t) + \sum_{j=1}^{t} m_1(t-j) f(j). \end{aligned} \tag{S4}$$

替换 $m_1(t) = \text{coverage@cost}(t)$ 即可得到公式（6）。类似地，我们现在可以得到二阶矩。

我们使用公式（S2）中的递归关系，并将 $x^2(t) \overset{d}{=} (1 + x(t-j))^2 = 1 + 2x(t-j) + x^2(t-j)$ 代入时间 $j \le t$。

$$\begin{aligned} &m_2(t) = \langle x^2(t) \rangle = \langle \langle x^2(t) \mid T^{(1)} \rangle \rangle = \sum_{j=1}^{\infty} \langle [x^2(t) \mid T^{(1)} = j] \rangle f(j) = \\ &\sum_{j=1}^{\infty} \mathbb{I}_{\{j \le t\}} \left(1 + 2\langle x(t-j) \rangle + \langle x^2(t-j) \rangle\right) f(j) = \sum_{j=1}^{t} (1 + 2m_1(t-j) + m_2(t-j)) f(j) = \\ &\sum_{j=1}^{t} (2 + 2m_1(t-j) - 1 + m_2(t-j)) f(j) = 2m_1(t) - F(t) + \sum_{j=1}^{t} m_2(t-j) f(j). \end{aligned} \tag{S5}$$

总的来说，我们得到

$$m_2(t) = 2m_1(t) - F(t) + \sum_{j=1}^{t} m_2(t-j) f(j). \tag{S6}$$

A.2. 式（7）的证明

首先，我们对方程（ Z）进行变换S3和（S6），得到

$$\begin{aligned}\mathcal{Z}\{m_1\}(z) &= \frac{\widetilde{f}(z)}{(1-z)\left(1-\widetilde{f}(z)\right)},\\ \mathcal{Z}\{m_2\}(z) &= \frac{\widetilde{f}(z)\left(1+\widetilde{f}(z)\right)}{(1-z)\left(1-\widetilde{f}(z)\right)^2}.\end{aligned} \tag{S7}$$

~其中 $f(z) \equiv \sum_{t=0}^{\infty} f(t)z^t$是回答问题所需时间的概率质量函数的 Z-变换。我们~ ~利用了累积分布函数的 Z-变换与概率质量函数之间的关系， $F(z) = f(z)/(1-z)$，以及 Z-变换的卷积定理（Feller，1991）。

取长时间极限等价于取极限 $z \to 1$，并对变换进行反演。所有反演均使用 Z-变换的陶伯定理完成（Feller，1991）。如果 $1-\text{pass@}k$ 比 ck^{-1}衰减得更快，且 c 为~某个常数，那么回答问题的平均时间将为 $\mathbb{E}[T]$，且 $f(z) = 1-(1-z)\mathbb{E}[T] + o(1-z)$。因此，式（ $z=1$）在 S7附近的渐近行为为

$$\begin{aligned}\mathcal{Z}\{m_1\}(z) &\simeq \frac{1}{(1-z)^2\mathbb{E}[T]} \Rightarrow m_1(t) \simeq \frac{t}{\mathbb{E}[T]},\\ \mathcal{Z}\{m_2\}(z) &\simeq \frac{2}{(1-z)^3\mathbb{E}^2[T]} \Rightarrow m_2(t) \simeq \frac{t^2}{\mathbb{E}^2[T]}.\end{aligned} \tag{S8}$$

我们得到，在这种情况下，coverage@cost 随时间线性增长，其斜率是回答单个问题所需平均尝试次数的倒数。

同样明显的是，已回答的独特问题数量的标准差 $\sqrt{m_2(t) - m_1^2(t)}$ 在长时间下的增长速度慢于 $\overline{t}$。因此，它与 coverage@cost 的比值在后期趋近于 0。这意味着在该情形下，$x(t)/m_1(t)$ 变为确定性的，而 coverage@cost 能够很好地描述其动力学。

另一种情况是 $1-\text{pass@}k$ 按 $ck^{-\alpha}/\Gamma(1-\alpha)$ 衰减，且满足 $0 < \alpha < 1$，对此我们得到，在 $\sim z = 1$ 附近， $f(z) = 1 - c(1-z)^{\alpha} + o((1-z)^{\alpha})$。于是

$$\begin{aligned}\mathcal{Z}\{m_1\}(z) &\simeq \frac{1}{c(1-z)^{1+\alpha}} \Rightarrow m_1(t) \simeq \frac{t^{\alpha}}{c\Gamma(1+\alpha)},\\ \mathcal{Z}\{m_2\}(z) &\simeq \frac{2}{c^2(1-z)^{1+2\alpha}} \Rightarrow m_2(t) \simeq \frac{2t^{2\alpha}}{c^2\Gamma(2\alpha+1)}.\end{aligned} \tag{S9}$$

可以看到，此时 coverage@cost 的行为变为次线性，其幂次由推理缩放幂 α 控制。

Regarding $\sigma_{x(t)}$ x（t）中的标准差及其与 coverage@cost 的比值；在此处，它们趋近于 h

$$\begin{aligned}\sigma_{x(t)} &\simeq \sqrt{\frac{2}{\Gamma(2\alpha+1)} - \frac{1}{\Gamma^2(\alpha+1)}}\frac{t^{\alpha}}{c},\\ \frac{\sigma_{x(t)}}{m_1(t)} &\simeq \sqrt{\frac{2\Gamma^2(\alpha+1)}{\Gamma(2\alpha+1)} - 1} \in (0,1).\end{aligned} \tag{S10}$$

此处，标准差与已回答不重复问题的平均数量之比趋近于一个常数。该常数的值随推理缩放幂的取值而变化，并且对于$0 < \alpha < 1$单调递减。

coverage@cost 的线性行为与次线性行为之间的转变出现在 $\alpha = 1$。当 $\alpha > 1$ 时，随机选取一个问题的平均回答时间有限；而当 $\alpha < 1$ 时，该时间发散。从数学上看，这与输运现象物理学中所用连续时间随机行走模型里的扩散行为和亚扩散行为之间的转变相同（Scher&Montroll，1975；Metzler & Klafter，2000）。

B. 定理 4.1 的证明

证明。 我们将证明 $\mathbb{E}[T_{\tau+1}] \geq \mathbb{E}[T_\tau]$ 对于 $\tau \geq 1$ 以及任意 $\mathcal{P}(p)$均成立。为此，我们定义如下差值$D(\tau)$:
$= \mathbb{E}[T_{\tau+1}] - \mathbb{E}[T_\tau]$并证明其非负。我们使用公式（11）来表示该差值

$$D(\tau) = \frac{G(\tau+1)}{F(\tau+1)} - \frac{G(\tau)}{F(\tau)} = \frac{F(\tau)G(\tau+1) - F(\tau+1)G(\tau)}{F(\tau)F(\tau+1)}. \tag{S11}$$

由于式（S11）中的分母非负，因此只需证明分子也非负。我们首先回顾

$$F(\tau) \coloneqq \Pr(T \leq \tau) = 1 - \int_0^1 (1-p)^\tau \mathcal{P}(p)\mathrm{d}p = 1 - \mathbb{E}\left[(1-p)^\tau\right], \tag{S12}$$

and

$$G(\tau) \coloneqq \sum_{k=0}^{\tau-1}(1-F(k)) = \int_0^1 \sum_{k=0}^{\tau-1}(1-p)^k \mathcal{P}(p)\mathrm{d}p = \int_0^1 \frac{1-(1-p)^\tau}{p}\mathcal{P}(p)\mathrm{d}p = \mathbb{E}\left[\frac{1-(1-p)^\tau}{p}\right]. \tag{S13}$$

因此，式（S11）的分子可写为

$$F(\tau)G(\tau+1) - F(\tau+1)G(\tau) = \mathbb{E}\left[1-q^\tau\right]\mathbb{E}\left[\frac{1-q^{\tau+1}}{p}\right] - \mathbb{E}\left[1-q^{\tau+1}\right]\mathbb{E}\left[\frac{1-q^\tau}{p}\right]. \tag{S14}$$

其中令 $q\colon = 1-p$。现在注意到 $\mathbb{E}\left[1-q^{\tau+1}\right] = \mathbb{E}\left[1-(1-p)q^\tau\right] = \mathbb{E}\left[1-q^\tau\right] + \mathbb{E}\left[pq^\tau\right]$，类似地$\mathbb{E}\left[\frac{1-q^{\tau+1}}{p}\right] = \mathbb{E}\left[\frac{1-(1-p)q^\tau}{p}\right] = \mathbb{E}\left[\frac{1-q^\tau}{p}\right] + \mathbb{E}\left[q^\tau\right]$。代入式（S14）右侧，得到

$$F(\tau)G(\tau+1) - F(\tau+1)G(\tau) = \mathbb{E}\left[1-q^\tau\right]\mathbb{E}\left[q^\tau\right] - \mathbb{E}\left[\frac{1-q^\tau}{p}\right]\mathbb{E}\left[pq^\tau\right]. \tag{S15}$$

现在定义 $u\colon = (1-q^\tau)/p$，并继续推导式（S15）右侧

$$\begin{aligned}&\mathbb{E}\left[1-q^\tau\right]\mathbb{E}\left[q^\tau\right] - \mathbb{E}\left[pq^\tau\right]\mathbb{E}[u] = \mathbb{E}[pu]\mathbb{E}[q^\tau] - \mathbb{E}[pq^\tau]\mathbb{E}[u] = \\ &\mathbb{E}[u]\mathbb{E}[q^\tau]\left(\frac{\mathbb{E}[pu]}{\mathbb{E}[u]} - \frac{\mathbb{E}[pq^\tau]}{\mathbb{E}[q^\tau]}\right) = \mathbb{E}[u]\mathbb{E}[q^\tau]\left(\frac{\mathbb{E}[pu]}{\mathbb{E}[u]} - \frac{\mathbb{E}[upq^\tau/u]\mathbb{E}[u]}{\mathbb{E}[u]\mathbb{E}[uq^\tau/u]}\right).\end{aligned} \tag{S16}$$

为完成证明，我们将 $\mathbb{E}_\lambda[p] = \mathbb{E}[\lambda(p)p]/\mathbb{E}[\lambda(p)]$ 定义为关于概率密度$\mathcal{P}(p)\lambda(p)/\mathbb{E}[\lambda(p)]$ 的期望，其中 $\lambda(p)$是正函数。于是，等式（S16）右侧的括号项可写为

$$\mathbb{E}_u[p] - \mathbb{E}_{q^\tau}[p] = \frac{\mathbb{E}_u[p]\mathbb{E}_u[q^\tau/u] - \mathbb{E}_u[pq^\tau/u]}{\mathbb{E}_u[q^\tau/u]} = -\frac{\mathrm{COV}_u(p, q^\tau/u)}{\mathbb{E}_u[q^\tau/u]} \geq 0.$$

最后一个不等式成立，是因为 $q^\tau/u \geq 0$，并且由于 $q^\tau/u = p(1-p)^\tau/(1-(1-p)^\tau)$ 随 p 严格递减，根据切比雪夫和不等式的连续版本，其协方差为负。因此， $D(\tau) \coloneqq \mathbb{E}[T_{\tau+1}] - \mathbb{E}[T_\tau] \geq 0$，证毕。

□

推论 B.1. 对于任意 $\mathcal{P}(p)$，以及 $\forall \tau \in \mathbb{N} \setminus \{0\}, \mathbb{E}[T] \geq \mathbb{E}[T_\tau]$。

证明。 回顾 $D(\tau) \geq 0$，并注意到在等式（11）中取极限 $\tau \to \infty$可得 $\lim_{\tau\to\infty} \mathbb{E}[T_\tau] = \mathbb{E}[T]$。因此，$\mathbb{E}[T] = \sup\{\mathbb{E}[T_\tau]\}$。

□

C. 评估 llama-3.1-8b-instant 的幂律指数

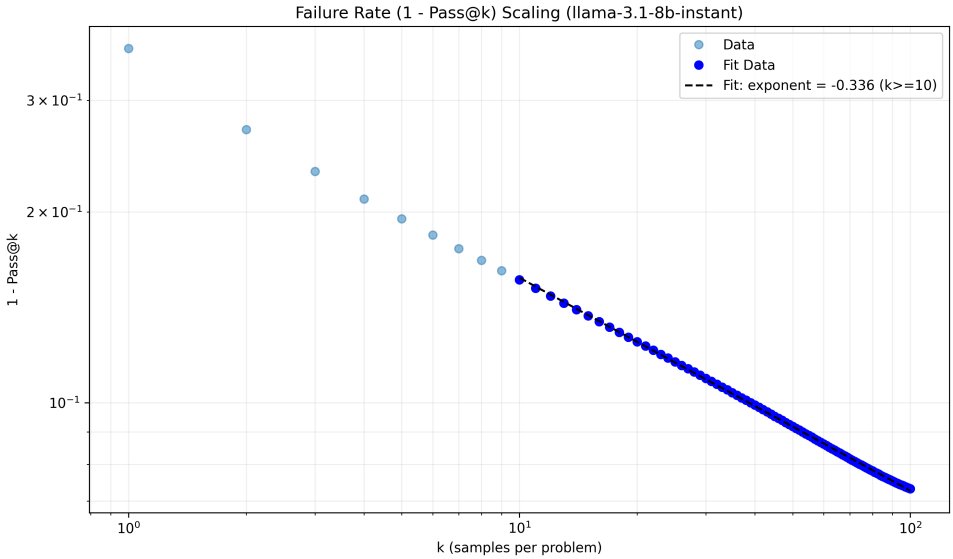


图 *S1*. llama-3.1-8b-instant 的 $\log(1-\text{pass@}k)$ 与 $\log k$ 的线性拟合。对于 $k \gg 1$，确实有 $(1-\text{pass@}k) \propto k^{-\alpha}$，得到的幂律指数为 $\alpha = 0.34$。