

The Crossover Does Not Carry the Claim: Pass@k Inversion, the Reasoning Boundary, and Five Premises the RLVR Debate Leaves Unstated

Pranay M. Mahendrakar

SONYTECH

ORCID: 0009-0003-7224-029X

Life of Research

2026-08-23

Abstract

A model trained with reinforcement learning from verifiable rewards beats its base model at one sample and loses to it at many. The field has read that crossover as a measurement of the reasoning boundary and has spent two years arguing about which direction the boundary moved. This paper separates the crossover, which is an observation reproduced across model families and benchmarks, from the boundary claim, which is an inference, and states the five premises the inference requires: that a large-k pass reflects reasoning rather than a hit in a small answer space, that a benchmark-averaged decline reflects a per-problem decline, that the training run being measured is not itself biased against the metric, that decoding and estimation were controlled across the two arms, and that the result generalises past static single-turn math. Every one of the five now has published counter-evidence, and none of that counter-evidence restores the optimistic reading either, because each rescue holds only under a condition its authors state and the optimistic reading does not. The paper then observes something the debate has not: six independent groups have since mid-2025 proposed an account under which the crossover is compatible with RLVR expanding what a model can solve, and the six accounts are not the same account. They locate the effect in different places - training phase, per-problem saturation, rollout-group sparsity, answer-space geometry, reasoning validity, task type - and they make different predictions. Pass@k cannot adjudicate between them, because it is the quantity all six agree is being misread. This paper reports no experiments. It states what the record establishes, states flatly what it does not, proposes a five-line disclosure that would make crossovers comparable, and names four studies that would settle the open part.

1. Introduction

The observation is not in dispute. Yue et al. (2025) sampled reinforcement learning with verifiable rewards (RLVR) models and their base models at increasing k across model families, RL algorithms, and math, coding and visual reasoning benchmarks, and reported that the trained models outperform their base models at small k while the base models achieve a higher pass@k score when k is large. The paper was given an oral at NeurIPS 2025. Its coverage and perplexity analyses concluded that "the observed reasoning abilities originate from and are bounded by the base model". A quantitative analysis in the paper, with the base model taken as an upper bound, found little to choose between six popular RLVR algorithms, none of which came close to making optimal use of the base model's potential. By contrast, the same paper reported that distillation can introduce new reasoning patterns from a teacher and genuinely expand what the model can do.

The crossover has since been reproduced or corroborated in enough settings that treating it as an artefact of one lab's pipeline is no longer available. Wu, F., et al. (2025) characterise RLVR as a support-constrained optimisation mechanism and report that the shrinkage of empirical support generally

outweighs its expansion under larger sampling budgets. Nguyen et al. (2025) report negative interference, in which learning to solve some training problems reduces the likelihood of correct solutions to others, together with a winner-take-all effect in which RLVR disproportionately reinforces problems whose correct solutions already have high likelihood under the base model. Ni et al. (2025) prove that Group Relative Policy Optimization is a conservative reweighting scheme bounded by the base model's distribution, and corroborate it on transformers trained from scratch. Fan, M., et al. (2026) formalise the same phenomenon as over-sharpening and attribute it to finite-batch updates biasing learning toward sampled modes. Saha et al. (2026) report that the entropy collapse in RLVR models is not merely stylistic: it is accompanied by a significant reduction in semantic branching entropy, measured on trees extracted from reasoning traces by semantic equivalence. The pattern is not confined to RLVR. Ge et al. (2026) report the same shape for on-policy distillation, and Strozzi (2026) reports it for teacher-free self-training on a domain with a free exact verifier.

Read off that list, the question looks settled. It is not, because a second list exists and is roughly as long. Liu, M., et al. (2025) report that prolonged RL training with KL control and reference-policy resetting uncovers reasoning strategies inaccessible to base models, including cases where the base model fails entirely regardless of the number of attempts. Wen et al. (2025) argue that pass@K credits answers rather than reasoning, introduce CoT-Pass@K, which requires the intermediate steps to be sound as well, and report that under that metric RLVR extends the boundary on math and code. Shen et al. (2025) built a synthetic maze environment in which difficulty is regulated by path length and number of turns, and report RLVR succeeding on problems where the base vision-language model attains zero percent accuracy however large the pass@k budget. Yuan, S., et al. (2026) report that a non-trivial fraction of initially unsolvable problems become solvable during ordinary RLVR training, and that restricting updates to problems with no observed success lifts Pass@256 above the base model on difficult benchmarks. Zhai et al. (2026) report that for agentic tool use the RL curve pulls above the base curve and the gap widens at large k rather than converging.

The natural move, and the one the field has made, is to say that both sides are right. This paper agrees that both sides are right and argues that saying so has concealed rather than resolved the problem, for two reasons.

The first is that the inference everyone is arguing about has never been written down. "RLVR-trained pass@k falls below base pass@k at large k" is a measurement. "RLVR does not expand the reasoning boundary" is a claim about a model's capability. Getting from the first to the second requires bridging premises, and because the premises are unstated they are also unaudited. Section 3 states five. Section 4 shows that each of the five now has published counter-evidence - and, symmetrically, that none of the counter-evidence licenses the optimistic conclusion either, because each rescue holds only under a scope condition its own authors state.

The second reason is more specific and, as far as the author can determine, has not been noted. Since mid-2025 at least six independent groups have proposed an account under which the crossover is compatible with RLVR expanding what a model can solve. They are not the same account. Yao et al. (2025) locate the split in training time, as two stages of a probability-mass dynamic. Yuan, S., et al. (2026) locate it in per-problem saturation, arguing that standard RLVR is structurally biased against high-k pass@k so that an aggregate decline is uninformative. Zhou (2026) locates it in rollout-group sparsity, as an absence-of-evidence failure on boundary prompts. Dragoi et al. (2025) locate it in the geometry of the answer space, arguing that at large k on discrete numeric answers pass@k measures the chance of a hit rather than reasoning. Wen et al. (2025) locate it in the validity of the reasoning that produced the answer. Zhai et al. (2026) locate it in task type, reporting that the pessimistic result holds for static reasoning and reverses for compositional tool use. These six accounts make different predictions and would be settled by different experiments. Section 5 sets them against each other. The

metric that would ordinarily arbitrate cannot, because it is the quantity all six say is being misread.

This paper is a position paper. It ran no experiments and reports no measurements of its own; everything quantitative in it is attributed. Its contribution is an audit of an inference, a demonstration that the field's apparent convergence hides six incompatible mechanisms, a five-line disclosure that would make published crossovers comparable to each other, and four studies that would settle the part that remains open. Section 6 argues the deepest of the confusions is not empirical at all: sharpening and expansion are being treated as exclusive alternatives when at instance level they are not.

2. The instrument, and what it is being asked to certify

Pass@k was not designed to measure a reasoning boundary. Chen, M., et al. (2021) introduced it to report functional correctness of generated code, with an unbiased estimator over n sampled completions, and used it to report that Codex solved 28.8 percent of HumanEval problems with one sample against 70.2 percent with one hundred. The interest of the metric there was practical: a developer can look at several candidate programs, and a test suite can filter them. Pass@k answered the question "if the user gets k tries and a verifier, how often do they succeed."

Repeated sampling then turned out to scale further than anyone expected. Brown et al. (2024) report that coverage, the fraction of problems solved by any sample, scales with the number of samples over four orders of magnitude, often log-linearly, and that on SWE-bench Lite the fraction of issues resolved by DeepSeek-Coder-V2-Instruct rises from 15.9 percent with one sample to 56 percent with 250, against a 43 percent single-sample state of the art. The same paper records the limit of the technique: where answers cannot be checked automatically, majority voting and reward models plateau beyond several hundred samples and fail to use the budget. The plateau is worth holding on to, because majority voting over sampled chains - self-consistency, in the formulation of Wang, X., et al. (2022) - is the aggregation rule a reader is most likely to have in mind when told that a model "can" solve a problem, and it is the rule that stops improving first.

Between that result and the arrival of RLVR at scale (Shao, Z., et al., 2024; DeepSeek-AI et al., 2025), pass@k acquired a second job. It stopped being a statement about a deployment with k tries and became a proxy for what a model can in principle do - the "reasoning boundary." The usage is now routine enough to appear without argument: Liang et al. (2025), for instance, describe Pass@k as the quantity that "typically represents the upper bound of LLM reasoning capability." Once that identification is made, the crossover reads as a statement about capability rather than about sampling, and the RLVR debate follows.

It is worth being precise about what the identification assumes, because the two jobs are different in kind. As a deployment statement, pass@k is definitionally correct: it is exactly the success rate of a k -try procedure with a verifier, and nothing needs to be true of the model's internals for the number to mean what it says. As a capability statement, pass@k is a proxy, and a proxy inherits every confound that separates the proxy from the target. The move from the first to the second happens silently, usually in the transition from a results table to a discussion paragraph.

Three properties of pass@k at large k make the proxy reading fragile, and all three are now documented rather than conjectural.

It is not monotone in what one would want a boundary to track. Yu (2025) derives the gradient of the pass@k objective and shows that it is a per-example positive reweighting of the pass@1 objective, that it provides a vanishing learning signal precisely where exploration matters most, and that pass@k and pass@1 draw closer together the more tightly the policy concentrates its probability mass. The metric's sensitivity therefore depends on the very property - dispersion - that the training procedure is changing.

It is statistically unstable at the sample counts people can afford. Hariri et al. (2025) report that $\text{pass}@k$ produces unstable and potentially misleading rankings when the number of trials is limited, and propose a Bayesian posterior with credible intervals in its place; they note that under a uniform prior the posterior mean is order-equivalent to average accuracy, which is the $\text{pass}@1$ regime, not the large- k one. Liu, J., et al. (2024) reach a compatible conclusion from a different direction with G-Pass@k, which reports stability alongside potential. Hochlehnert et al. (2025) report that math reasoning benchmarks are highly sensitive to decoding parameters, random seeds, prompt formatting and even hardware and software configuration, and that under a standardised protocol most RL approaches yield only modest improvements, far below prior claims, with small benchmarks such as AIME'24 the worst offenders. Miller (2024) had already set out the statistical machinery this implies.

It counts hits, not reasons. This is the property that does the most damage, and Section 4.1 treats it separately.

A third job has since been added on top of the second. Chen, Z., et al. (2025) use Pass@k directly as the training reward, derive an analytical solution for the resulting advantage, and report improved exploration, concluding that exploration and exploitation are not inherently conflicting; Thrampoulidis et al. (2025) show that direct REINFORCE-style methods for the Pass@K objective and advantage-shaping modifications of GRPO are two sides of one surrogate-reward formulation. Yu (2025) argues from the same gradient that $\text{pass}@k$ is a useful diagnostic and an unsuitable objective. Whatever the merits of optimising it, the promotion matters here for a narrow reason: a quantity being simultaneously used as the field's capability definition, its evaluation metric and its training signal is a quantity under considerable load.

None of this shows that the crossover is an artefact. It shows that the instrument is doing a job it was not built for, in a regime where its behaviour is known to be unstable, and that the field has been reading a difference of two such measurements as a difference of two capabilities.

3. The inference, stated as five premises

Write the observation as O and the claim as C.

O: on benchmark B, with decoding configuration D, estimated from n samples per problem, the RLVR-trained policy attains a lower $\text{pass}@k$ than the base model for k above some crossover point.

C: RLVR did not expand the model's reasoning boundary; it redistributed probability mass within a support the base model already had.

C does not follow from O alone. It follows from O together with the following, each of which is an empirical claim in its own right and none of which is routinely reported.

P1 (answer-channel validity). A counted pass at large k reflects a reasoning process the model executed, rather than a hit in a small answer space or a correct answer reached by an unsound derivation.

P2 (aggregation faithfulness). A decline in the benchmark-averaged $\text{pass}@k$ reflects a net contraction of the per-problem solvable set, and not a larger contraction on some problems masking a real expansion on others.

P3 (optimisation neutrality). The particular RLVR run being measured is not itself structurally biased against high- k $\text{pass}@k$, such that the decline is a property of the training configuration rather than of RLVR as a family.

P4 (comparison control). The base and trained arms were compared under matched decoding and matched estimation, so that the crossover is not partly an artefact of temperature, prompt format,

sample count or estimator variance.

P5 (scope). The result generalises from the setting where it was measured - overwhelmingly static, single-turn, verifier-checkable math on a small number of model families - to the claim about reasoning capability that is being made.

These are stated as premises of the pessimistic inference because that is the inference the field is arguing about. Each has a mirror image on the optimistic side, and Section 4 states those too. The point of the exercise is not that the pessimistic reading is wrong. It is that O plus five unexamined premises has been reported as though it were O alone.

4. Auditing the five

4.1 P1: what a large-k pass actually certifies

Two independent failures separate "a sample was scored correct" from "the model solved the problem," and both worsen with k .

The first is the answer space. Dragoi et al. (2025) point out that when answers come from a small discrete set - a numeric result to a math problem is the typical case - the chance that some sample lands on the right one keeps growing with the number of samples, so $\text{pass}@k$ at large k comes to measure the number of attempts rather than reasoning, and reading it as a reasoning boundary can mislead. Their alternative, $\text{Cover}@tau$, credits a problem only when a fraction tau or more of its completions are right; a model that succeeds by lucky hits loses ground quickly as tau is raised, and they report that the ordering of popular algorithms shifts relative to $\text{Pass}@1$. The direction of the bias matters here and is usually left implicit: a higher-entropy policy accumulates hits faster than a concentrated one, so the confound systematically favours the base model in exactly the comparison the crossover is used to make.

The second is the derivation. Wang, Y., et al. (2025) examined false positive solutions - correct final answers reached through flawed deduction - across models, datasets and decoding strategies, and report that they persist throughout, that sampling-based inference-time scaling does not alleviate the problem, and that $\text{pass}@N$ is more susceptible to them than single-sample evaluation, implying a scaling ceiling significantly lower than automatic evaluation indicates. Yuan, Y., et al. (2025) give the failure a taxonomy, naming Miracle Steps, abrupt jumps to a correct output with no valid preceding derivation, and link them by probing to answer-recall shortcuts including pretraining memorisation. Training against a rubric reward model rather than answer matching raised their Verified $\text{Pass}@1024$ on AIME2024 from 26.7 percent to 62.6 percent and reduced Miracle Steps by 71 percent.

Wen et al. (2025) build the corresponding metric and reach the corresponding conclusion: scoring a sample only when both the answer and the intermediate steps are sound, they report that RLVR extends the reasoning boundary for mathematics and code, and offer a theoretical account of why answer-only rewards would nonetheless incentivise correct reasoning.

What this establishes is narrower than it first appears, and the narrowing runs against this paper's own argument as much as against the pessimistic one. The false-positive literature shows the aggregate credit assigned by $\text{pass}@k$ at large k is inflated for both arms. It does not by itself show the inflation is asymmetric. Dragoi et al. give an argument that it should be, by way of dispersion; Wen et al. give a measurement under a metric whose reasoning-validity judgement is itself model-mediated, which imports a second instrument with its own error rate. The honest statement is that P1 is false as a universal claim, that its failure is documented rather than suspected, and that nobody has published a decomposition of how much of any specific crossover it accounts for. That decomposition is Study 1 in Section 8.

4.2 P2: the aggregate is a net of two flows

The benchmark-averaged $\text{pass}@k$ curve is a sum over problems, and the papers that have looked underneath it report movement in both directions at once.

Yuan, S., et al. (2026) report, as an observational finding, that over the course of standard RLVR training a non-trivial share of the problems that were unsolvable at the outset become solvable. Wu, F., et al. (2025) state their own headline result as a comparison of two terms rather than the absence of one: "the shrinkage of empirical support generally outweighs the expansion of empirical support under larger sampling budgets". That is a claim about a difference. The expansion term is not zero in their data; it is smaller. Ge et al. (2026) make the two flows explicit for on-policy distillation, running a problem-level solvability analysis with $\text{pass}@1024$ as the criterion and reporting an asymmetry in which more previously solvable problems become unsolvable than the reverse. Nguyen et al. (2025) supply a mechanism for the losing flow in negative interference and winner-take-all dynamics, and show it is addressable: curating training data toward low-likelihood problems improves $\text{pass}@k$.

The consequence for the inference is direct. C is a claim about the support of the policy. O is a claim about an average. An average that declines is consistent with a solvable set that has both grown and shrunk, and the papers that report the decomposition find exactly that. A crossover therefore establishes a net, and the net is the least informative summary of the two numbers that produced it.

The limit of this objection should be stated as plainly. A net contraction is still a contraction, and if the deliverable is a deployed model then the net is what the user gets. P2's failure does not rescue the optimistic reading; it shows that the pessimistic reading is answering a different question - about the artefact - than the one it is usually taken to answer, about the method. It also does not tell us how large the expansion term is, because only three of the papers surveyed here report it at all, and Ge et al.'s number is for on-policy distillation rather than RLVR.

4.3 P3: the run is configured against the metric

This is the sharpest of the five, and it is the least widely absorbed.

Yuan, S., et al. (2026) explain the decline as overtraining. After a problem has contributed all it can to the reference metric, more updates on it add nothing to the set of problems the model can solve; they only pile more probability onto whichever trajectories on-policy sampling already prefers. With the usual handful of rollouts per problem, one success is enough to put a problem in that saturated state as far as high- k $\text{pass}@k$ is concerned. On their account, then, most of what standard RLVR does is overtraining when judged by the boundary, and the method as normally configured leans structurally against high- k $\text{pass}@k$. Their interventional result is the one that matters for P3: when updates are confined to problems on which no success has yet been observed, $\text{Pass}@256$ rises above the base model on difficult benchmarks.

Zhou (2026) reaches a compatible diagnosis from rollout statistics rather than saturation, arguing that it is on boundary prompts that the failure is concentrated: the base model holds correct trajectories for them that sampling can recover, but these are rare, too sparse to turn up reliably in finite RLVR rollout groups, so the group never sees what it would need to reinforce. Liu, M., et al. (2025) reach it from training duration, reporting that prolonged training with KL control and reference-policy resets uncovers strategies inaccessible to base models even under extensive sampling, with boundary improvement correlating with base-model competence and training duration. Dong et al. (2025) reach it from the on-policy restriction itself, countering what they call capability boundary collapse by admitting external data through multiple importance sampling. Liang et al. (2025) reach it from problem supply, reporting that synthesising variational problems from the policy's own correct solutions maintains entropy and improves $\text{Pass}@32$ by 18.3 and 22.8 absolute points on AIME24 and AIME25

over standard RLVR. Cui et al. (2025) supply the underlying dynamic: policy entropy collapses early and predictably, they fit $R = -a \exp(H) + b$ between entropy and performance, and they note the ceiling this implies at $H = 0$. Deng et al. (2025) survey the exploration mechanisms involved across training stages, instances and tokens, and treat the entropy-performance exchange as the quantity a configuration trades against.

Taken together these results indicate that the crossover is a property of a particular training configuration - on-policy, few rollouts per problem, no entropy management, no external data, stopped early - rather than of RLVR as a family. The strength of that indication needs stating precisely, because the papers are not all claiming the same thing. Three of the modifications above are reported to lift the trained policy above the base model at large k : restricting updates to unsolved problems (Yuan, S., et al., 2026), prolonged training with KL control and reference resets (Liu, M., et al., 2025), and admitting external data (Dong et al., 2025). Two more are reported only to improve on standard RLVR, with no base-model comparison at large k stated in the same breath: problem synthesis (Liang et al., 2025) and per-prompt base anchoring (Zhou, 2026). The second pair is the weaker result and should not be counted as a reversal. Even on the first three, a statement about RLVR as a family cannot be supported by measurements of one configuration of it, however widely that configuration is used.

Two things restrain this conclusion. First, every one of the interventions listed is a proposal by the authors who report the effect, evaluated by those authors; none has an independent replication in the corpus surveyed here. Second, an intervention that raises $\text{pass}@k$ may do so by raising dispersion, and if P1 fails then raising dispersion raises the hit rate whether or not it raises capability. P1 and P3 interact, and no paper in this corpus reports them jointly.

4.4 P4: matched decoding, matched estimation

The two arms of a crossover are two models evaluated under some configuration, and the configuration is doing more work than it is credited with.

Wu, Y., et al. (2025) report that at large K further scaling yields no gains and some hard questions remain unsolved regardless of the number of traces, that different sampling temperatures solve different subsets of problems, and that scaling along the temperature dimension yields an additional 7.3 points averaged over Qwen3 at 0.6B, 1.7B, 4B and 8B and five benchmarks. The finding with direct bearing on P4 is their report that temperature scaling enables base models to reach performance comparable to RL-trained counterparts without any post-training. If a decoding change can close the gap the training was measured to produce, then a single-temperature comparison between a base model and a trained one is not a comparison of two capabilities. It is a comparison of two capability-and-decoding pairs, and the trained arm has usually had its decoding tuned for $\text{pass}@1$ while the base arm has not had its decoding tuned at all.

Hochlehnert et al. (2025) add the estimation half: seeds, prompt format, hardware and software configuration all move these numbers, and small benchmarks move most. Hariri et al. (2025) add that $\text{pass}@k$ is the least stable of the available summaries at achievable sample counts.

The restraint here is real and cuts the other way too. Yue et al. (2025) sampled across families, algorithms and modalities, and a decoding artefact that survives that much variation is not a plausible sole explanation of the crossover. P4's failure is best read not as an alternative explanation but as a missing error bar: the published crossovers are point comparisons whose sensitivity to configuration has, with the exception of Hochlehnert et al., not been characterised.

4.5 P5: what the setting was

The crossover literature is overwhelmingly static, single-turn, numeric-answer mathematics, with code as a secondary domain, on a small number of open model families. Three results indicate the conclusion

does not transfer cleanly out of that setting.

Shen et al. (2025) constructed synthetic maze navigation with difficulty controlled by path length and number of turns, and report RLVR succeeding on problems where the base vision-language model attains zero percent accuracy despite increasing pass@k budgets, with zero-shot gains on two real navigation benchmarks. A base model at zero percent at every sampled k cannot be the observed source of the trained model's solutions, which makes this, in the corpus surveyed here, the most direct counter-example to the strong bounded-by-the-base reading - in one synthetic domain, on one model family.

Zhai et al. (2026) introduce PASS@(k,T), varying sampling budget and interaction depth jointly, and report that for agentic tool use the RL curve pulls above the base curve with the gap widening at large k rather than converging; that the expansion is specific to compositional sequential information gathering and disappears on simpler tasks; and that supervised fine-tuning on matched data regresses the boundary on the same tasks. Their own summary is that the optimistic and pessimistic readings are both correct on different task types.

Shao, R., et al. (2025) attack the generality of the model sample rather than the task sample. Training Qwen2.5-Math-7B with GRPO on randomly assigned rewards improved MATH-500 by 21.4 percentage points, against 29.1 points from ground-truth rewards; they attribute this to a clipping bias amplifying high-prior pretraining behaviours, identify code reasoning as one such behaviour rising from 65 percent to over 90 percent frequency, and report that spurious rewards effective on Qwen models often fail on Llama3 or OLMo2. If a large fraction of a measured RLVR effect on one family is amplification of a pretraining behaviour rather than learning, then conclusions about what RLVR can and cannot do drawn from that family are conclusions about that family.

The scope restraint runs both ways here, and the optimistic papers need it more. Shen et al.'s environment is synthetic and its difficulty axis is stipulated by construction. Zhai et al.'s expansion is explicitly confined to compositional tool use and explicitly absent on simpler tasks. Liu, M., et al.'s ProRL result is on a 1.5B model. None of these establishes that the crossover in static math is wrong; each establishes that the conclusion drawn from it does not travel.

5. Six reconciliations that are not the same reconciliation

The field's current position is that both readings are partly right. That position is correct and is also where the analysis has stopped. Set the proposals side by side and they are incompatible in their mechanism, in their predictions, and in what would refute them.

Yao et al. (2025) locate the split in training time. On their account RLVR runs through two stages of probability-mass dynamics: an exploitation stage in which positive advantage raises already-high-reward tokens and suppresses low-reward ones while leaving a potentially optimal token's probability largely unchanged, and an exploration stage, reached only after the first saturates, in which an occasionally sampled optimal token starts to gain. Shrinkage is what over-exploitation looks like; expansion requires getting into stage two. The prediction is temporal: run longer and the sign flips.

Yuan, S., et al. (2026) locate it in per-problem saturation. Shrinkage is not a stage the run passes through but a property of which problems the updates land on, and the aggregate decline is uninformative because the metric is structurally biased. The prediction is compositional: change the problem selection and the sign flips without running longer.

Zhou (2026) locates it in rollout-group sparsity. The loss is concentrated on boundary prompts whose correct trajectories are too rare to appear in a finite group. The prediction is about group size and prompt regime: raise the rollout budget on the right prompts, or anchor them to the base distribution, and the losses do not occur.

Dragoi et al. (2025) locate it in the answer space. Part of the base model's large-k advantage is not solving but hitting, and the effect is a property of discrete numeric answers, not of RLVR. The prediction is about the task: move to a task with a large or continuous answer space, or raise the reliability threshold, and the base model's advantage shrinks or disappears.

Wen et al. (2025) locate it in reasoning validity. The base model's extra passes at large k include a higher share of unsound derivations. The prediction is about the grader: score reasoning as well as answers and the sign flips.

Zhai et al. (2026) locate it in task type. Static reasoning behaves as the pessimists say and compositional tool use behaves as the optimists say. The prediction is about the domain, and specifically about whether the task rewards multi-step interaction that resampling cannot recover.

These are not layers of one explanation. Yao et al. and Yuan, S., et al. disagree about whether the shrinkage phase is necessary or avoidable. Zhou and Yuan, S., et al. both point at problem-level structure but attribute the loss to different quantities - sampling sparsity versus advantage saturation - which imply different fixes. Dragoi et al. and Wen et al. both say the measurement is wrong but disagree about which part of it is wrong, and their corrections are not composable in an obvious way: Cover@tau makes the reliability threshold explicit and would penalise a model that reasons correctly but rarely, while CoT-Pass@K makes the derivation explicit and would credit exactly that model. Zhai et al.'s reconciliation says the others are all describing one task family.

There is a further problem that applies to all six. Each was validated by showing that its proposed intervention improves pass@k, or a variant of pass@k, relative to a baseline. If the metric is the thing under suspicion, then improving it is not evidence that the mechanism named is the operative one. Several of these accounts would predict the same pass@k improvement from different causes, and the corpus surveyed here contains no study that pits two of them against each other on a shared benchmark with a shared protocol. That absence, rather than the disagreement, is the reason the debate has not closed.

6. Sharpening and expansion are not exclusive

Underneath the empirical disagreement is a conceptual one that no amount of measurement will settle, because it concerns what the two words mean.

The debate is conducted as a disjunction: RLVR either expands the reasoning boundary or merely sharpens the distribution over what the base model could already produce. The disjunction is treated as exhaustive and exclusive. Wang, Z., et al. (2026) give the argument for why it is neither. Defining capability by instance-level solvability, they observe that success on a multi-step problem is governed by the joint probability of its atomic steps, which decays exponentially in the number of steps. Sharpening the probability of each atomic step therefore raises the joint probability multiplicatively, and can lift an instance from effectively unreachable to reliably solvable without any new primitive being acquired. Training models exclusively on single-step operations in the Algebrarium framework and evaluating on unseen multi-step tasks, they report Pearson correlations between composite performance and the joint probability of atomic steps in the range 0.69 to 0.96, and report that RLVR acting as a global optimiser can sacrifice specific skills to maximise aggregate reward.

A compatible observation comes from a very different setup. Zhao, R., et al. (2025) trained models from scratch on open data mixtures and compared PPO, GRPO and Expert Iteration; they report that RL algorithms converge toward a dominant output distribution that amplifies patterns present in the pretraining data, and, in the same study, that RL post-training on simpler questions can produce gains on harder ones. Those two findings are usually quoted separately and read as belonging to opposite camps. On the instance-level account they are one finding: amplification of existing patterns is what

produces the transfer to harder problems.

If that is right, then "merely sharpening" and "expanding what the model can solve" are not alternatives. Sharpening is a mechanism by which the solvable set grows. The two camps have been arguing about a dichotomy that dissolves once capability is defined at the instance level rather than at the level of a support over trajectories, and the disagreement is partly about which definition is in play. Yue et al.'s coverage and perplexity analyses concern the support: whether the trained model's solutions were reachable under the base distribution. Shen et al.'s zero-percent-at-every-k result concerns instance solvability: whether the base model in practice ever produces one. A trajectory can be in the support and have probability low enough that no achievable k recovers it, in which case the support claim is true and the practical claim is false. Wu, F., et al. (2025) come close to naming this when they report that seemingly more uncertain paths converge onto a smaller set of distinct answers, which is a statement about the gap between token-level and answer-level dispersion.

Mukherjee et al. (2025) supply a fact that fits either reading and should therefore be handled carefully. Across seven RL algorithms and ten models they report that RL updates only 5 to 30 percent of parameters, that finetuning the subnetwork alone recovers test accuracy and produces a nearly identical model, and that subnetworks overlap across seeds, data and algorithms more than chance. Read as support for sharpening, this is a small edit to an existing computation. Read against it, sparsity of the update says nothing about the size of the behavioural change, since the updates are reported as nearly full-rank. The result constrains the mechanism and does not decide the question, and it is cited here as an example of evidence both camps could claim.

7. What the record establishes, and what it does not

Stated flatly, so that the calibration is visible.

Established. The crossover is real and reproducible: across model families, RLVR algorithms, and math, code and visual benchmarks, RLVR-trained policies exceed their base models at small k and are exceeded by them at large k (Yue et al., 2025), with corroboration under independent framings (Wu, F., et al., 2025; Nguyen et al., 2025; Fan, M., et al., 2026) and analogues outside RLVR (Ge et al., 2026; Strozzi, 2026). Policy entropy collapses early and monotonically under standard configurations, and performance saturates with it (Cui et al., 2025), with the collapse extending to semantic and not merely stylistic diversity (Saha et al., 2026). Under GRPO specifically, and under their formalisation of it, Ni et al. (2025) prove the reweighting is bounded by the base distribution.

Reported but not independently replicated. The crossover is not a fixed property of RLVR. Three groups report a configuration change that puts the trained policy above its base model at large k : restricting updates to unsolved problems (Yuan, S., et al., 2026), prolonging training with KL control and reference resets (Liu, M., et al., 2025), and admitting off-policy data (Dong et al., 2025). Two further changes are reported to improve large- k performance relative to standard RLVR without a stated base-model comparison (Liang et al., 2025; Zhou, 2026), and entropy management is reported to raise downstream performance without a crossover claim attached (Cui et al., 2025). Every one of these reports is by the group proposing the change, and none has an independent replication in this corpus. This is the weakest link in the present argument and Section 9 says so.

Established. Pass@ k at large k is a poor instrument for the job it is being put to. It credits unsound derivations, and is more susceptible to them than single-sample evaluation (Wang, Y., et al., 2025), it credits hits in small answer spaces (Dragoi et al., 2025), it ranks unstably at achievable sample counts (Hariri et al., 2025; Liu, J., et al., 2024), its sensitivity depends on the dispersion the training is changing (Yu, 2025), and the benchmarks it is computed on are sensitive to configuration details that are rarely reported (Hochlehnert et al., 2025).

Not established, and this is the flat statement the calibration table asks for. Nobody has published a decomposition of a single crossover into its components: how much is guessing, how much is unsound derivation, how much is per-problem contraction net of expansion, how much is the training configuration, how much is decoding. In the absence of that decomposition the magnitude of the real effect is unknown, and both the strong pessimistic claim and the strong optimistic claim are underdetermined by the published record. Nobody has run two of the six reconciliations against each other. Nobody has reported a crossover with a matched-decoding control arm. And no operational definition of "reasoning boundary" has been proposed that all parties would accept and that any achievable measurement could estimate; the term is currently defined by the metric used to measure it, which is why the disagreement has been so hard to resolve.

8. A disclosure, and four studies

8.1 Five lines that would make crossovers comparable

A paper reporting a pass@k crossover could add five lines that cost nothing to produce and would let a reader locate the result against the five premises. None of the papers surveyed here reports all five.

Line one, the estimation. The number of samples drawn per problem, the estimator used, and a confidence or credible interval at the largest k reported. A point estimate at $k = 256$ taken from 256 samples is not comparable to one taken from 1024, and at present a reader usually cannot tell which they are looking at.

Line two, the decoding, for both arms separately. Temperature, top-p and prompt template for the base arm and for the trained arm, with a note on whether either was tuned. Given Wu, Y., et al. (2025), an untuned base arm compared against a trained arm tuned for pass@1 is not a controlled comparison, and the direction of the resulting bias is not obvious.

Line three, the two flows. The number of problems that moved from unsolved to solved and from solved to unsolved at the largest k, and not only the net of the two. Ge et al. (2026) demonstrate that this costs one extra pass over per-problem records that the evaluation has already produced.

Line four, the saturation profile. Rollouts per problem during training, and the fraction of training problems with at least one success in the rollout group. That fraction is what determines whether the overtraining regime of Yuan, S., et al. (2026) applies to the run being reported.

Line five, the answer space. For each benchmark, whether answers are numeric, multiple-choice or open, and the success rate a uniform guess would achieve. This is what bounds the size of the confound identified by Dragoi et al. (2025).

8.2 Four studies

Study 1: decompose one crossover. Take a single well-documented crossover and partition the large-k passes into sound derivations and false positives using a validated grader, and separately into hits explicable by answer-space size. Report the crossover with each component removed. This is assembling existing instruments - the false-positive taxonomy of Yuan, Y., et al. (2025), the CoT-Pass@K construction of Wen et al. (2025), the Cover@tau construction of Dragoi et al. (2025) - on one benchmark rather than proposing a new one. Until it is done, the size of the effect under dispute is unknown.

Study 2: adjudicate two reconciliations. Yao et al. (2025) predict the sign flips with training duration at fixed problem selection; Yuan, S., et al. (2026) predict it flips with problem selection at fixed duration. These are separable in a two-by-two design on one model and one benchmark. The corpus surveyed here contains no experiment that varies both.

Study 3: the matched-decoding crossover. Re-run a published crossover with both arms tuned independently for large-k coverage, including temperature scaling in the sense of Wu, Y., et al. (2025). Report whether the crossover point moves and by how much. This is the cheapest of the four and would settle P4 for at least one setting.

Study 4: instance solvability versus support. Section 6's dichotomy is testable. On a task family where atomic steps can be isolated - the Algebrarium setting of Wang, Z., et al. (2026) is one candidate, the maze construction of Shen et al. (2025) another - measure whether instances the trained model solves were reachable under the base distribution at any achievable k, and report the two definitions of capability separately. A finding that they systematically disagree would explain much of the debate without anyone having been wrong.

9. What would falsify this reading

The position argued here is that the crossover is a valid observation whose standard interpretation rests on five unaudited premises, and that the field's apparent convergence on "both are right" conceals six incompatible mechanisms. It would be substantially wrong under any of the following.

If a decomposition along the lines of Study 1 found that guessing and false positives together account for a small fraction of the crossover at large k - say under ten percent of the gap - then P1's failure would be real but immaterial, and the pessimistic reading would be stronger than this paper allows.

If the interventions of Section 4.3 failed to replicate independently, then P3's counter-evidence would collapse, and the crossover would look like a property of RLVR rather than of a configuration. This is the most likely of the three, since none of the interventions currently has an independent replication.

If the six reconciliations turned out to be reformulations of one dynamic - plausibly, if per-problem saturation, rollout sparsity and the two-stage account are three descriptions of the same advantage-estimation effect at different granularities - then Section 5's central claim would be a taxonomy error rather than a finding. The author regards this as the most serious threat to the paper's contribution, and notes that Thrampoulidis et al. (2025) have already shown two apparently distinct families of pass@k policy-gradient methods to be two sides of one surrogate-reward objective, which is precisely this kind of collapse occurring in an adjacent literature.

10. Limitations of this evidence base

This paper is an argument over a literature and inherits that literature's weaknesses.

The corpus is arXiv-centric and English-only. It was assembled by targeted search on the terms of the debate, not by a systematic protocol with stated inclusion criteria and a screened candidate pool, so a paper that makes the argument of Section 3 or Section 5 in different vocabulary could have been missed. The claim that the six reconciliations have not been set against each other is a claim about what the author found and is scoped accordingly.

Several load-bearing sources are recent preprints without refereeing or independent replication. Zhou (2026) is a single-author preprint, and Sections 1, 4.3, 5 and 7 use it for the rollout-sparsity account. Yuan, S., et al. (2026), whose overtraining argument carries much of Section 4.3 and is the strongest single result in this paper's case, was posted in June 2026. Wang, Z., et al. (2026), Fan, M., et al. (2026), Saha et al. (2026), Ge et al. (2026), Zhai et al. (2026) and Strozzi (2026) are all 2026 preprints. Strozzi (2026) states its own limitation - four trajectories, described by the author as indicative rather than a robust interval - and is cited here only as an instance of the crossover shape appearing outside RLVR. If the 2026 material does not hold up, Sections 4.2, 4.3 and 5 weaken together.

The paper takes abstracts and reported results at face value. Every citation was verified to exist and to

match its title and authors, and every attributed number was read against the source's own statement of it, but the author did not re-analyse anyone's data and cannot vouch for any of the measurements.

Finally, the audit is asymmetric by construction. Section 3 states the premises of the pessimistic inference because that is the inference under dispute. The optimistic inference has its own unstated premises - most importantly, that an intervention which raises pass@k has raised capability rather than dispersion - and Sections 4.3 and 5 note this but do not develop it. A companion audit of the optimistic side is owed and is not attempted here.

11. Conclusion

Two years of argument about whether reinforcement learning from verifiable rewards expands or narrows what a language model can reason about have been conducted through a single number, taken from a code-generation paper that used it to describe a k-try deployment, and pressed into service as the operational definition of a capability. The crossover that number exhibits is real. The claim it is asked to carry requires five further things to be true, and each of the five now has published counter-evidence: large-k passes include hits and unsound derivations, benchmark averages hide per-problem expansion, the standard training configuration is biased against the metric, the two arms are not decoding-matched, and the setting is narrower than the conclusion.

The field has reached the position that both readings are partly right. Six papers have said so and have meant six different things by it, locating the effect in training phase, problem saturation, rollout sparsity, answer-space geometry, reasoning validity and task type. Those accounts predict different outcomes and are not currently distinguished by any published experiment, because the metric that would distinguish them is the one they all say is being misread. Deciding among them does not require a new benchmark or a new algorithm. It requires reporting the five lines in Section 8.1 alongside the crossover, and running the four studies in Section 8.2 with instruments the literature has already built.

Acknowledgements

This paper is a reading of other people's experiments and owes its entire substance to the groups that ran them and published enough detail to be argued with. The authors of the papers whose premises are audited here made that audit possible by stating their scope conditions plainly; the disagreements described above are disagreements about inference, not about candour.

Revision note

Version 1.1 (29 September 2026) replaces version 1.0. Passages that followed the wording of cited abstracts too closely are now either quoted, with their citation, or rewritten in new words, and a few qualifiers the sources state (such as "may" and "often") were restored in the process. The argument, the references and the reported values are unchanged.

References

- Brown, B., Juravsky, J., Ehrlich, R., Clark, R., Le, Q. V., Re, C., & Mirhoseini, A. (2024). Large Language Monkeys: Scaling Inference Compute with Repeated Sampling. arXiv:2407.21787. [PREPRINT]
- Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P. de O., Kaplan, J., Edwards, H., Burda, Y., Joseph, N., Brockman, G., et al. (2021). Evaluating Large Language Models Trained on Code. arXiv:2107.03374. [PREPRINT]
- Chen, Z., Qin, X., Wu, Y., Ling, Y., Ye, Q., Zhao, W. X., & Shi, G. (2025). Pass@k Training for Adaptively Balancing Exploration and Exploitation of Large Reasoning Models. arXiv:2508.10751. [PREPRINT]
- Cui, G., Zhang, Y., Chen, J., Yuan, L., Wang, Z., Zuo, Y., Li, H., Fan, Y., Chen, H., Chen, W., Liu, Z., Peng, H., Bai, L., Ouyang, W., Cheng, Y., Zhou, B., & Ding, N. (2025). The Entropy Mechanism of Reinforcement Learning for

Reasoning Language Models. arXiv:2505.22617. [PREPRINT]

DeepSeek-AI, Guo, D., Yang, D., Zhang, H., Song, J., Wang, P., Zhu, Q., Xu, R., Zhang, R., Ma, S., et al. (2025). DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. *Nature*, 645, 633-638. doi:10.1038/s41586-025-09422-z; arXiv:2501.12948.

Deng, J., Chen, J., Chen, Z., Cheng, D., Bai, F., Zhang, B., Min, Y., Gao, Y., Zhao, W. X., & Wen, J.-R. (2025). From Trial-and-Error to Improvement: A Systematic Analysis of LLM Exploration Mechanisms in RLVR. arXiv:2508.07534. [PREPRINT]

Dong, Y., Jiang, X., Tao, Y., Liu, H., Zhang, K., Mou, L., Cao, R., Ma, Y., Chen, J., Li, B., Jin, Z., Huang, F., Li, Y., & Li, G. (2025). RL-PLUS: Countering Capability Boundary Collapse of LLMs in Reinforcement Learning with Hybrid-policy Optimization. *ACL 2026 (main)*. arXiv:2508.00222.

Dragoi, M., Pintilie, I., Gogianu, F., & Brad, F. (2025). Beyond Pass@k: Breadth-Depth Metrics for Reasoning Boundaries. arXiv:2510.08325. [PREPRINT]

Fan, M., Han, W., Wang, D., Chen, C., Zhang, Z., & Zhou, J. (2026). When Sharpening Becomes Collapse: Sampling Bias and Semantic Coupling in RL with Verifiable Rewards. arXiv:2601.15609. [PREPRINT]

Ge, X., Zhang, Z., Huang, Y., Zhu, J., Yuan, L., Gu, W., Wu, W., Huang, W., Zhang, X., Han, B., Zhou, J., & Yao, J. (2026). Towards Understanding On-Policy Distillation through the Lens of Test-Time Scaling. arXiv:2608.11829. [PREPRINT]

Hariri, M., Samandar, A., Hinczewski, M., & Chaudhary, V. (2025). Don't Pass@k: A Bayesian Framework for Large Language Model Evaluation. *ICLR 2026*. arXiv:2510.04265.

Hochlehnert, A., Bhatnagar, H., Udandaraao, V., Albanie, S., Prabhu, A., & Bethge, M. (2025). A Sober Look at Progress in Language Model Reasoning: Pitfalls and Paths to Reproducibility. *COLM 2025*. arXiv:2504.07086.

Liang, X., Li, Z., Gong, Y., Shen, Y., Wu, Y. N., Guo, Z., & Chen, W. (2025). Beyond Pass@1: Self-Play with Variational Problem Synthesis Sustains RLVR. arXiv:2508.14029. [PREPRINT]

Liu, J., Liu, H., Xiao, L., Wang, Z., Liu, K., Gao, S., Zhang, W., Zhang, S., & Chen, K. (2024). Are Your LLMs Capable of Stable Reasoning? *ACL 2025*. arXiv:2412.13147.

Liu, M., Diao, S., Lu, X., Hu, J., Dong, X., Choi, Y., Kautz, J., & Dong, Y. (2025). ProRL: Prolonged Reinforcement Learning Expands Reasoning Boundaries in Large Language Models. arXiv:2505.24864. [PREPRINT]

Miller, E. (2024). Adding Error Bars to Evals: A Statistical Approach to Language Model Evaluations. arXiv:2411.00640. [PREPRINT]

Mukherjee, S., Yuan, L., Hakkani-Tur, D., & Peng, H. (2025). Reinforcement Learning Finetunes Small Subnetworks in Large Language Models. *NeurIPS 2025*. arXiv:2505.11711.

Nguyen, P. M., La, C. D., Nguyen, D. M. H., Chawla, N. V., Nguyen, B. T., & Doan, K. D. (2025). The Reasoning Boundary Paradox: How Reinforcement Learning Constrains Language Models. arXiv:2510.02230. [PREPRINT]

Ni, K., Tan, Z., Liu, Z., Li, P., & Chen, T. (2025). Can GRPO Help LLMs Transcend Their Pretraining Origin? arXiv:2510.15990. [PREPRINT]

Saha, S., Sharma, K., Chaturvedi, A., & Asher, N. (2026). BODHI: Do LLMs Branch Out and Discover Heterogeneous Inferences? arXiv:2608.02867. [PREPRINT]

Shao, R., Li, S. S., Xin, R., Geng, S., Wang, Y., Oh, S., Du, S. S., Lambert, N., Min, S., Krishna, R., Tsvetkov, Y., Hajishirzi, H., Koh, P. W., & Zettlemoyer, L. (2025). Spurious Rewards: Rethinking Training Signals in RLVR. arXiv:2506.10947. [PREPRINT]

Shao, Z., Wang, P., Zhu, Q., Xu, R., Song, J., Bi, X., Zhang, H., Zhang, M., Li, Y. K., Wu, Y., & Guo, D. (2024). DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models. arXiv:2402.03300. [PREPRINT]

Shen, M., Zhi, Z., Liu, C., Xing, S., Tu, Z., & Liu, C. (2025). Does RLVR Extend Reasoning Boundaries? Investigating Capability Expansion in Vision-Language Models. arXiv:2511.00710. [PREPRINT]

Strozzi, I. L. (2026). Teacher-Free Self-Training Amplifies but Does Not Compound: A Pass@K Crossover on a Free-Verifier Domain. arXiv:2606.07856. [PREPRINT - single author, not refereed]

Thramopoulos, C., Mahdavi, S., & Deng, W. (2025). Advantage Shaping as Surrogate Reward Maximization: Unifying Pass@K Policy Gradients. *Transactions on Machine Learning Research*. arXiv:2510.23049.

- Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A., & Zhou, D. (2022). Self-Consistency Improves Chain of Thought Reasoning in Language Models. ICLR 2023. arXiv:2203.11171.
- Wang, Y., Yang, N., Wang, L., Wei, F., & Feng, F. (2025). Examining False Positives under Inference Scaling for Mathematical Reasoning. arXiv:2502.06217. [PREPRINT]
- Wang, Z., Li, Y., Zhang, S., Wang, Z., Zhang, H., Qu, X., & Cheng, Y. (2026). New Skills or Sharper Primitives? A Probabilistic Perspective on the Emergence of Reasoning in RLVR. arXiv:2602.08281. [PREPRINT]
- Wen, X., Liu, Z., Zheng, S., Ye, S., Wu, Z., Wang, Y., Xu, Z., Liang, X., Li, J., Miao, Z., Bian, J., & Yang, M. (2025). Reinforcement Learning with Verifiable Rewards Implicitly Incentivizes Correct Reasoning in Base LLMs. arXiv:2506.14245. [PREPRINT]
- Wu, F., Xuan, W., Lu, X., Liu, M., Dong, Y., Harchaoui, Z., & Choi, Y. (2025). The Invisible Leash: Why RLVR May or May Not Escape Its Origin. arXiv:2507.14843. [PREPRINT]
- Wu, Y., Mirhoseini, A., & Tambe, T. (2025). On the Role of Temperature Sampling in Test-Time Scaling. arXiv:2510.02611. [PREPRINT]
- Yao, X., Yu, L., Hu, X., Teng, F., Cui, Q., Zhou, J., & Liu, Y. (2025). The Debate on RLVR Reasoning Capability Boundary: Shrinkage, Expansion, or Both? A Two-Stage Dynamic View. arXiv:2510.04028. [PREPRINT]
- Yu, Y. (2025). Pass@k Metric for RLVR: A Diagnostic Tool of Exploration, But Not an Objective. arXiv:2511.16231. [PREPRINT - single author]
- Yuan, S., Chen, J., Zheng, J., Li, M., Feng, L., Wang, D., Xiang, T., Liu, T., & An, B. (2026). Understanding Diversity Collapse in RLVR via the Lens of Overtraining. arXiv:2606.15455. [PREPRINT]
- Yuan, Y., Mang, Q., Chen, J., Wan, H., Liu, X., Xu, J., Huang, J., Wang, W., Jiao, W., & He, P. (2025). Curing Miracle Steps in LLM Mathematical Reasoning with Rubric Rewards. ACL 2026 (main). arXiv:2510.07774.
- Yue, Y., Chen, Z., Lu, R., Zhao, A., Wang, Z., Yue, Y., Song, S., & Huang, G. (2025). Does Reinforcement Learning Really Incentivize Reasoning Capacity in LLMs Beyond the Base Model? NeurIPS 2025 (Oral). arXiv:2504.13837.
- Zhai, Z., Yan, W., Shao, X., & Wang, X. (2026). Does RL Expand the Capability Boundary of LLM Agents? A PASS@k(T) Analysis. arXiv:2604.14877. [PREPRINT]
- Zhao, R., Metere, A., Kakade, S., Pehlevan, C., Jelassi, S., & Malach, E. (2025). Echo Chamber: RL Post-training Amplifies Behaviors Learned in Pretraining. COLM 2025. arXiv:2504.07912.
- Zhou, T. (2026). When RLVR Shrinks the Reasoning Boundary: Diagnosing Pass@k Inversion. arXiv:2607.20543. [PREPRINT - single author, not refereed]

(c) 2026 Pranay Mahendrakar. Licensed under CC BY 4.0.

The literature search, drafting and citation verification for this paper were carried out with AI assistance under the author's direction. Every citation was machine-verified against the arXiv API and Crossref before inclusion, and every quantitative claim was read back against the cited source's own text. The author is responsible for the final text and for all claims made in it.

DOI: 10.5281/zenodo.22063127