POWERFLOW · ICML 2026 · UNSUPERVISED CAPABILITY ELICITATION

One actor, two roles.

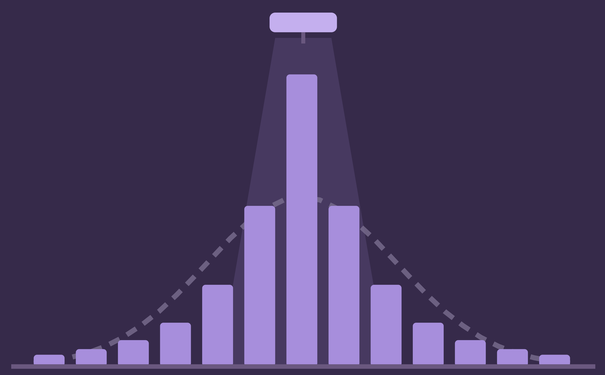
Unlocking the dual nature of LLMs by matching the base model's *own* α-power distribution — one control knob re-directs the same actor: **sharpen** to reason, **flatten** to create. No labels, no external rewards — no new cast.

Ruishuo Chen · Yu Chen · Zhuoran Li · Longbo Huang✉ — Institute for Interdisciplinary Information Sciences, Tsinghua University

ON STAGE · A IS THE DIRECTOR'S NOTE

Same actor, one note — a tight beam or a full-stage wash.

A > 1 · SPOTLIGHT

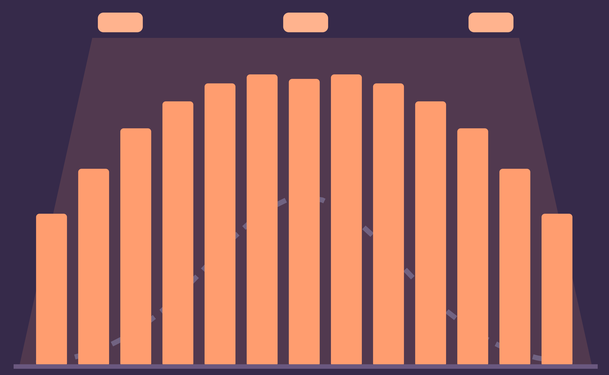


α = 4

Sharpen → reasoning

One tight beam: the stage's whole energy gathers onto the strongest reasoning paths.

A < 1 · ENSEMBLE WASH



α = 0.5

Flatten → creativity

The wash opens the full stage: voices muted by alignment return as creative diversity.

Director's note. A principled re-direction, not a reward hack: the length-aware objective is the **minimum-distortion** (I-projection) length correction, and on a trained model ~**91%** of the base model's mode rankings survive (inversion rate ≈ 0.09). Same actor, re-directed — never re-cast.

THE PREMISE · SAME ACTOR

The cast never changes — only the direction does.

PowerFlow recasts unsupervised fine-tuning as **distribution matching**: it targets the base model's own α-power (escort) distribution — no labels, no external rewards, no verifiers.

$$p_\alpha(y|q) \propto p_{\text{base}}(y|q)^\alpha$$

The single knob α re-directs it while preserving the base model's relative mode rankings: **α>1** concentrates mass on latent reasoning paths; **α<1** reopens the creative modes alignment suppresses.

A *length-aware* Trajectory-Balance objective cancels autoregressive length bias — so sharpening no longer collapses into trivially short outputs, nor flattening into repetition.

SPOTLIGHT · A > 1 · REASONING

Sharpen, and the reasoning line steps forward.

avg@16 over six math & science benchmarks · α=4 (base), α=2 (instruct)

42.17 Qwen2.5-Math-7B average

- Tops **every unsupervised method** on this model (TTRL 41.18, EMPO 40.88, PowerSampling 39.80) and is on par with *supervised* GRPO (42.38).
- Beats supervised GRPO on **3 of 4** tested models, comparable on the 4th; all claimed gains exceed 1σ.
- Solution diversity on AIME24/25 **4.05** vs GRPO 3.93, EMPO 3.80: sharper, not narrower.

ENSEMBLE WASH · A < 1 · CREATIVITY

Flatten, and the whole ensemble speaks again.

α=0.5 · three creative genres × four instruct models · diversity, redundancy (ROUGE-L↓), LLM-judged quality

The **only** evaluated method to raise *both* diversity and quality over the Instruct model — a Pareto-dominant shift that keeps instruction-following intact.

`Qwen2.5-1.5B-Instruct, overall`
`diversity 0.229 → 0.331  redundancy 0.270 → 0.113  quality 0.495 → 0.514`

PowerFlow · ICML 2026 · Ruishuo Chen, Yu Chen, Zhuoran Li, Longbo Huang✉ (corresponding)

Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China · Code `github.com/Chenruishuo/PowerFlow` · longbohuang@tsinghua.edu.cn

CODE