NeurIPS 2026

# Reasoning as Compression

## Unifying Budget Forcing via the Conditional Information Bottleneck

Fabio Valerio Massoli  Andrey Kuzmin  Arash Behboodi

Qualcomm AI Research†

Qualcomm

NEURAL INFORMATION PROCESSING SYSTEMS

† Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

# Budget forcing

## The problem

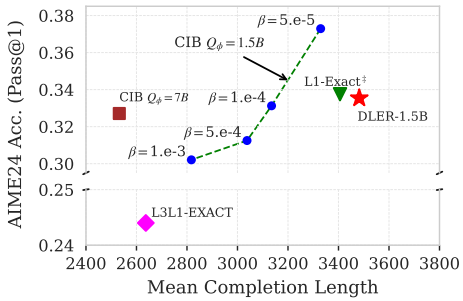
- Chain-of-Thought (CoT) prompting improves LLM accuracy tasks but often increases token usage and inference cost.
- Existing "Budget Forcing" methods reduce cost via fine-tuning with heuristic length penalties, suppressing both essential reasoning and redundant filler.

## Our approach

Reasoning as **lossy compression**: with the **Conditional Information Bottleneck (CIB)** each token pays its *surprisal* under a frozen LM.

## Contributions

- **Theory:** the IB fails for transformers; CIB fixes it.
- **Unification:** length penalties are special cases.
- **Results:** higher accuracy *and* shorter CoTs.




AIME24, DLER-1.5B: CIB (blue, brown) reaches higher accuracy with shorter CoTs than the baseline (red star); $\beta$ moves along the frontier.

# The Attention–Markov Incompatibility

**Information Bottleneck (IB)** (Tishby et al., 1999) seeks a representation $Z$ that maximally compresses the input $X$ while preserving information about the target $Y$:

$$\mathcal{L}_{\text{IB}} = I(X;Z) - \mu\, I(Y;Z). \tag{1}$$

Crucially, the standard IB assumes the Markov chain $Y \leftrightarrow X \leftrightarrow Z$, implying that $Z$ is the sole channel through which information flows from $X$ to $Y$.

**Attention breaks the Markov property**

Due to the causal attention mechanism, the decoder predicting $Y$ attends to **both** the prompt $X$ and the generated chain $Z$. This forms a collider structure: $(X, Z) \to Y$.

**Fix: the Conditional Information Bottleneck (CIB)**

View the prompt $X$ as *side information* that is always available to the answer generator; $Z$ encodes only the *additional* information necessary to predict $Y$ given $X$:

$$\mathcal{L}_{\text{CIB}} = I(X;Z) \;-\; \mu\, I(Y;Z \mid X). \tag{2}$$

# The Conditional Information Bottleneck

$$\max_{\theta} \underbrace{I(Z;Y \mid X)}_{\text{sufficiency}} - \beta \underbrace{I(X;Z)}_{\text{minimality}}$$

$Z \sim \pi_\theta(\cdot \mid X), Y \sim \pi_\theta(\cdot \mid X, Z)$: keep what $Z$ adds about $Y$ *beyond* $X$, pay for what $Z$ copies from $X$.

**Sufficiency (lower bound)**, $Q_\rho$ = answer verifier:

$$I(Z;Y|X) \gtrsim \mathbb{E}_{Z\sim\pi_\theta(\cdot|X)}\big[\log Q_\rho(y \mid x, Z)\big]$$

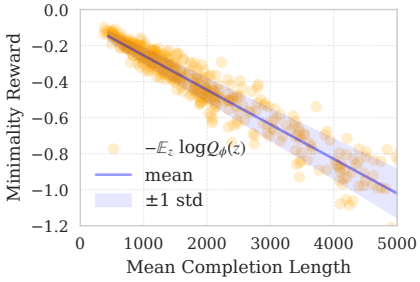
**Minimality (upper bound)**, $Q_\phi$ = frozen base LM:

$$I(X;Z) \le \mathbb{E}\big[-\log Q_\phi(Z)\big] - H(Z \mid X)$$

**CIB reward (trained with GRPO)**

$$R = \underbrace{\mathbb{1}\big[\hat{y}(x,z) = y\big]}_{r_{\text{acc}}} + \beta \underbrace{\textstyle\sum_t \log Q_\phi(z_t \mid z_{<t})}_{r_{\text{min}}}$$

Each token pays its **surprisal**; $Q_\phi$ is *not* used at inference.




$r_{\text{min}}$ falls with length, but its spread at fixed length shows the cost depends on *content*, not only on $|Z|$.

# Length penalties are CIB with uninformative priors

**Key property.** CIB serves as a general framework from which, e.g., length-based penalties naturally arise as a special case.
Each budget-forcing method is a choice of prior in the CIB minimality term.

## Prop. 1: linear length penalty ⇔ uniform prior

Max-entropy prior $Q(z_t) = 1/|V|$: every token has surprisal $\log |V|$,

$$-\log Q(Z) = -\sum_{t=1}^{T} \log \tfrac{1}{|V|} = T \log |V|.$$

so the CIB penalty is $\beta T \log |V| = \alpha T$, a linear length penalty.

**Implication:** all tokens carry the same information, regardless of meaning.

## Prop. 2: LCPO-Exact ⇔ Laplace prior

LCPO-Exact (L1; Aggarwal & Welleck, 2025) uses $|n_{\text{gold}} - T|$ as penalty:

$$Q_{\text{LCPO}}(Z) \propto e^{-|n_{\text{gold}} - T|}.$$

a Laplace prior over the length, centred at $n_{\text{gold}}$.

**Implication:** assumes a *golden length* exists.

Both priors depend **only on** $T = |Z|$. The CIB prior $Q_\phi(z_t \mid z_{<t})$ is **per-token and content-dependent**: predictable filler is cheap to drop, informative steps are kept if they raise accuracy.

# Results: better accuracy *and* shorter CoTs

| Model | Math | | | | | Science, Knowledge & Coding | | |
|---|---|---|---|---|---|---|---|---|
| | MATH500 | AIME24 | AIME25 | Minerva | Olympiad | GPQA-D | MMLU-Pro | LCB |
| *DeepScaleR-1.5B* | | | | | | | | |
| Baseline | 85.8 / 3190 | 38.2 / 9331 | 25.4 / 8907 | 19.9 / 6386 | 54.8 / 5922 | 36.5 / 7445 | 39.4 / 5285 | 18.7 / 11403 |
| L3L1-EXACT | 85.2 / 1932 | **24.4 / 2637** | **20.2 / 2608** | 19.8 / 2124 | **48.8 / 2369** | 37.9 / 3125 | 40.4 / 2240 | 17.5 / **4612** |
| ThinkPrune-I2k† | 85.5 / 1707 | 34.9 / 5095 | – | – | 54.7 / 3498 | – | – | – |
| LAPO-I† | 86.3 / 2168 | 38.1 / 5371 | – | – | **56.3** / 4024 | – | – | – |
| CIB$_{\beta^-,1.5B}$ | **87.6** / 2359 | **40.2** / 7298 | **27.8** / 6866 | **20.4** / 4510 | 55.5 / 4472 | 38.4 / 4571 | **41.6** / 3286 | **18.8** / 10410 |
| CIB$_{\beta^+,1.5B}$ | 87.2 / 2017 | 38.2 / 6937 | 27.6 / 6757 | 20.1 / 4147 | 55.3 / 4211 | 35.5 / 4429 | 40.7 / 3108 | 18.6 / 10274 |
| CIB$_{7B}$ | 86.1 / **1337** | 37.5 / 5003 | 25.0 / 5389 | 19.7 / **2058** | 55.0 / 3065 | **39.5 / 2850** | 40.8 / **2118** | 17.4 / 9481 |
| *Nemotron-1.5B* | | | | | | | | |
| Baseline* | 85.8 / 8509 | 43.1 / 22931 | **35.1** / 24086 | 11.0 / 13011 | 43.8 / 16270 | 31.4 / 21238 | 41.4 / 10490 | 26.6 / 24082 |
| L1-Exact‡ | 85.8 / 8807 | 40.2 / 22079 | 30.8 / 23243 | 11.4 / 12925 | 44.0 / 15316 | 33.9 / 21132 | 39.6 / 10416 | **26.8** / 23995 |
| CIB$_{\beta^-,1.5B}$ | **86.8** / 7536 | **45.3** / 21962 | 34.1 / 22980 | **11.7** / 11118 | **52.7** / 15089 | **34.9** / 22558 | 39.4 / 9851 | 26.1 / 23557 |
| CIB$_{\beta^+,1.5B}$ | 85.4 / 5982 | 41.5 / 20954 | 34.0 / 21407 | 10.9 / 10051 | 50.8 / 13894 | 34.8 / 19694 | 39.6 / 8908 | 26.1 / 22227 |
| CIB$_{7B}$ | 86.7 / **5789** | 42.7 / **20547** | 32.3 / **20678** | 10.6 / **8752** | 51.1 / **12242** | 33.6 / **19533** | **41.6 / 8496** | 25.9 / **21817** |
| *DLER-1.5B* | | | | | | | | |
| Baseline | 86.4 / 1873 | 33.4 / 3478 | 23.9 / 3290 | 20.9 / 2587 | 54.1 / 2766 | **40.9** / 3280 | 41.7 / 2575 | 17.9 / 4857 |
| L1-Exact‡ | 86.6 / 1751 | 35.0 / 3274 | 22.7 / 3037 | 21.2 / 2277 | 54.0 / 2577 | 36.8 / 3012 | 40.8 / 2326 | 17.8 / 4585 |
| CIB$_{\beta^-,1.5B}$ | **88.4** / 1687 | **35.4** / 3198 | **25.3** / 3260 | **21.6** / 2346 | **54.3** / 2616 | 39.4 / 3183 | **41.8** / 2469 | **18.2** / 4727 |
| CIB$_{\beta^+,1.5B}$ | 87.6 / 1421 | 34.1 / 3124 | 24.2 / 3027 | 21.0 / 2305 | 54.1 / 2544 | 38.8 / 3044 | 41.6 / 2353 | 17.9 / 4629 |
| CIB$_{7B}$ | 85.0 / **805** | 31.6 / **2504** | 22.4 / **2257** | 20.4 / **1327** | 51.6 / **1631** | 36.7 / **2423** | 40.4 / **1973** | 16.8 / **3977** |
| *DLER-7B* | | | | | | | | |
| Baseline | 93.6 / 1159 | 49.6 / 3221 | 36.6 / 3271 | 27.3 / 1849 | **65.0** / 2840 | 50.1 / 3255 | **58.7** / 2012 | **35.9** / 4847 |
| L1-Exact‡ | 92.8 / **675** | 46.2 / 2982 | 31.2 / 2564 | 26.1 / 1144 | 61.9 / 1872 | 48.6 / 2551 | 56.9 / 1446 | 34.4 / 4120 |
| CIB$_{\beta^-,7B}$ | **94.0** / 1033 | **49.7** / 2642 | **37.1** / 2962 | **27.6** / 1381 | 64.1 / 2361 | **50.5** / 2899 | **58.7** / 1718 | 35.6 / 4003 |
| CIB$_{\beta^+,7B}$ | 92.2 / 678 | 48.2 / **2419** | 35.4 / **2550** | 25.8 / **851** | 62.7 / **1704** | 49.7 / **2476** | 58.2 / **1430** | 34.5 / **3796** |

Acc. (%, pass@1, 16 samples) / avg. CoT tokens. **Bold**: best acc. / shortest CoT in group (**green**); **red**: significant drop. $\beta^- = 5\times10^{-5}$, $\beta^+ = 1.5\times10^{-4}$, subscript = prior size. † Wu et al., 2025; ‡ our L1-Exact runs; * pass@1 with $n=16$. CIB trained on math only.

## Cost of the prior

**Training:** a frozen 7B semantic prior is only $<$**6.2%** slower.

1.47 $\rightarrow$ 1.38 samples/s (DLER-1.5B, 4$\times$H100): one forward pass, no gradients.

**Inference:** **zero** extra cost; the prior is used only during training.

# Why a *semantic* prior? It changes what is kept

**Uniform vs. semantic prior.** Tuned uniform-prior baseline; compared at *matched length* ($\pm 1\%$):

| | Uniform | CIB | | Δ |
|---|---|---|---|---|
| | Acc. / Tok. | Setting | Acc. / Tok. | Acc. (↑) |
| DeepScaleR | 46.0 / 5224 | $\beta^-$ | 49.9 / 5113 | **+3.9** |
| DeepScaleR | 46.0 / 5224 | $\beta^+$ | 48.8 / 4870 | **+2.8** |
| DLER-1.5B | 47.0 / 2813 | $\beta^-$ | 48.6 / 2789 | **+1.6** |

**Redundancy markers** ("wait", "Alternatively", …); analysed 1432 paired completions:

- per completion: $-47\%$ (DeepScaleR), $-61\%$ (DLER)
- per 1k tokens: $-39\%$, $-50\%$. Markers become *rarer*, not just fewer.

The semantic cost changes the *composition* of the reasoning, not only its length.

**Prompt**

Two sides of a triangle are each 8 units long. If the third side has a whole number length, what is the greatest possible perimeter?

**Baseline (standard CoT)**

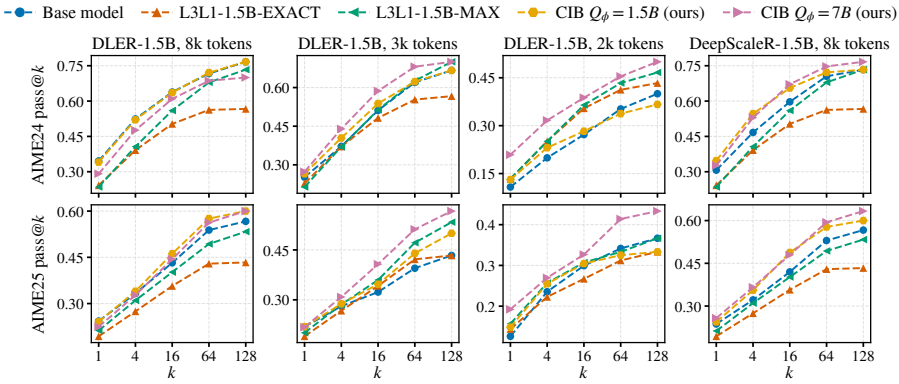
Okay, so I have this problem… Hmm, let me think about how to approach this. **Triangle inequality**: $8 + 8 > c$, so $c \leq 15$. But wait, wait. Let me think again… $8 + c > 8$ simplifies to $c > 0$, always true… But wait, let me double-check. Plug $c = 15$… Yes… Yes… Wait, but hold on. Let me make sure I didn't miss any constraints… Perimeter $= 8 + 8 + 15 =$ **31**.

**Ours: CIB**

Apply the **triangle inequality**; let the third side be $x$. (1) $8 + 8 > x \Rightarrow x < 16$; (2) $8 + x > 8 \Rightarrow x > 0$ (always true). Largest integer below 16 is 15: perimeter $8 + 8 + 15 =$ **31**.

CIB keeps the logic and drops the **redundant loops** (red).

# Scaling under token budgets




**Pass@$k$ on AIME24/25.** Columns 1–3: DLER-1.5B at 8K / 3K / 2K max tokens; column 4: DeepScaleR-1.5B at 8K.

- CIB scales with $k$ as well as or better than the base model, with shorter traces.
- At lower token busgets (2K and 3K), the 7B-prior shows the largest gains over the baselines.

## Conclusions and takeaways

- **CIB solves Attention–Markov Incompatibility.** In attention-based architectures, the the prompt $X$ is attended to directly during answer generation. Therefore it represents side information. The right objective is: $\max I(Z;Y \mid X) - \beta I(X;Z)$.
- **Length penalties are CIB with uninformative priors.** A uniform prior gives the linear length penalty; a Laplace prior gives LCPO-Exact. Budget forcing becomes a choice of prior.
- **A semantic prior prices information, not tokens.** Predictable filler is cheap to remove while informative steps are kept: better accuracy at shorter CoTs, beating a tuned uniform prior at matched length and cutting redundant self-verification loops.
- **The gains transfer.** Trained on math only, CIB stays on par or improves on GPQA, MMLU-Pro and LiveCodeBench with fewer tokens, and scales well with pass@$k$ under tight budgets.
- **A general recipe.** Pick a sufficiency signal and a prior: RLVR, RLAIF, on-policy distillation and SFT are all special cases.