ELF-REG: Scaling Continuous Diffusion Language Models to Reasoning Tasks

Better intermediate representations, useful answers before the trajectory ends

Paper authors: Zeyu Michael Li, William Xingxu Chen, Bingshuo Qian, Jiayin Liu, Xiang Cheng

Duke University / Tsinghua University

arXiv:2609.29102v3, 8 October 2026
Reading discussion: 14 October 2026

Can continuous parallel generation learn reasoning?

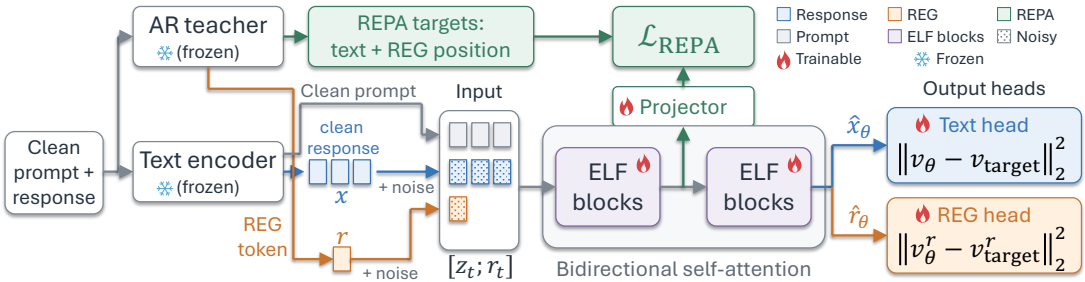
The question

Can a fully continuous language model produce useful mathematical solutions and code with few network function evaluations (NFE)?

- ELF (Embedded Language Flows) denoises all response positions together and discretizes only at the final decode.
- Training: align intermediate features and jointly denoise a global teacher representation.
- Inference: decode a clean prediction from an early trajectory prefix.

Paper, Secs. 1–3, PDF pp. 2–4. REPA, REG, and early stopping build on prior work.

Two teacher signals shape one bidirectional student



REPA: local alignment
Match intermediate student features to clean AR teacher states.

REG: a global generated state
Denoise one global representation together with the response.

Original Fig. 1(a), PDF p. 1; Sec. 3.2, p. 4. Teacher and projector are training-only.

ELF predicts clean states, then decodes once

$$z_t = tx + (1-t)\sigma\varepsilon, \quad \varepsilon \sim \mathcal{N}(0, I), \quad \sigma = 2$$

$$v_\theta = \frac{\hat{x}_\theta - z_t}{1-t}, \qquad z_{j+1} = z_j + \Delta_j v_j$$

- x: standardized, contextual text states from frozen Qwen3-0.6B-Base; width 1024.
- Prompt stays fixed; response starts from noise.
- The previous clean prediction conditions the next update.
- The shared Transformer also learns token decoding with cross-entropy; one final call emits all tokens.

Eq. (1), Sec. 3.1, PDF p. 4. Implementation clamps the velocity denominator at 0.05.

REPA supervises features; REG adds a reconstruction task

REPA

$$1 - \cos\big(p_\phi(u_i^{(\ell)}), h_i\big)$$

- Teacher: Qwen3-1.7B-Base, layer 20.
- Student block 4 (B) or 8 (L).
- Response content positions plus REG; student remains bidirectional.

REG

$$r_t = tr + (1 - t)\sigma\varepsilon_r$$

- r: standardized last content-token state, teacher layer 24.
- Joint attention and velocity regression.
- At inference, REG starts from noise and remains decoder context.

$$\mathcal{L} = \mathcal{L}_{\mathrm{ELF}} + 0.5\,\mathcal{L}_{\mathrm{REPA}} + 0.1\,\mathcal{L}_{\mathrm{REG}}$$

Eq. (2), PDF p. 4; Appendix A and Table 7, pp. 18–20.

Early-stop spends its budget on a longer grid's prefix

1. Draw a sorted logit-normal grid for master budget $M = 8N$.
2. Run only its first $N - 1$ denoiser evaluations.
3. Decode the latest predicted-clean state at decoder time 1.

Actual budget	Full-span	Prefix early-stop
64 NFE	63 updates over $[0, 1]$	First 63 updates of a 512-NFE grid
Decode input	Updated terminal state	Latest clean prediction

The observed exit time averages $t = 0.0825$ in the diagnostic runs.
$\rho = 8$ is an NFE ratio, not a continuous-time fraction.

Sec. 3.3, PDF p. 4; Algorithm 1, p. 27; Table 12, p. 29. NFE includes the decoder call.

Early-stop approximates the remaining flow map

$$F_{t,1}(y) = y + \int_t^1 v(y_s, s)\, ds, \qquad D_t(y) = y + (1-t)v(y,t)$$

$$e_t(y) = D_t(y) - F_{t,1}(y) = (1-t)\big(v(y,t) - \bar{v}_{t:1}(y)\big)$$

Endpoint-error bound

If v is C^1, the flow is unique through time 1, and $\|a_s\| \le A$ on $[t,1]$, then

$$\|e_t(y)\| \le \frac{A(1-t)^2}{2}.$$

Here y contains text and REG. An accurate early exit needs current velocity to approximate the remaining average velocity $\bar{v}_{t:1}$.

Proposition 1, Eq. (4), PDF p. 5; proof in Appendix C.1, p. 26.

Matching endpoint tokens is stronger than preserving an answer

$$L\|e_t(y_t)\| < m \quad \Longrightarrow \quad \mathcal{D}(\hat{y}_t) = \mathcal{D}(y_1)$$

- m: smallest winning-token logit margin through the first EOS.
- L: Lipschitz constant for *pairwise logit differences*.
- This sufficient condition preserves the whole endpoint response, including an incorrect one.

Three different questions

Are the latent states close? Are the token sequences identical?
Do the responses produce the same numerical answer?

Proposition 2 and Corollary 1, PDF p. 5; Appendix C.2, pp. 26–27.

The experiments train a separate model for each task

Task	Student	Main NFE	Training corpus
GSM8K	B / L	64	OpenMathInstruct-2 + MetaMathQA
MATH-500	L	128	OpenMathInstruct-2, after GSM8K
Python code	L	128	OpenCodeInstruct
MMLU	B	64	Auxiliary data + generated rationales

- B / L: 12 / 32 blocks; hidden width 768 / 1280.
- Main math/code tables: zero-shot, 16 samples per question, early-stop $\rho = 8$.
- pass@1 is mean single-sample accuracy; pass@k estimates whether at least one of k candidates succeeds.
- MMLU uses full-span generation. Generation seeds share trained weights.

Sec. 5, PDF p. 7; Tables 7–8, pp. 19–21; evaluation protocols, p. 22.

The larger model gains on both math benchmarks

Task	Size	NFE	ELF	ELF-REG	Gain (pp)
GSM8K	B	64	37.71	38.11	+0.40
GSM8K	L	64	51.85	55.96	+4.11
MATH-500	L	128	10.55	13.39	+2.84

- All values: pass@1 (%), averaged over 16 generation seeds.
- ELF-REG-B at 16 NFE: 34.42%; FMLM+ (Init) at 64 NFE: 33.60%.
- ELF-REG-L on GSM8K: 52.89% at 16 NFE, 55.04% at 32 NFE.
- B's final gap is small; its training-efficiency benefit is clearer.

Table 2, PDF p. 8; Table 14, p. 30. Cross-paper data and protocols differ.

Code gains appear across all four main benchmarks

Model	NFE	MBPP-500	MBPP-378	HumanEval	HE+
Open-dCoder	128	16.70	23.90	20.80	17.60
oDLM	128	15.70	22.99	17.87	16.40
PlaidQ+CFG	257	15.53	24.58	22.04	19.70
ELF-L	128	17.19	26.75	19.66	18.75
ELF-REG-L	128	**19.10**	**28.92**	**22.56**	**21.46**

- pass@1 (%); both MBPP variants use base tests.
- ELF uses EvalPlus sanitization and a single-function alias rule.
- ELF allows the remaining 1024-token canvas for its response; PlaidQ reports a 128-token completion cap.

Table 1, PDF p. 7; Appendix B.3 and B.5, pp. 22–25. HE+ = HumanEval+.

Function-name handling changes the MBPP ranking

ELF-REG-L, 128 NFE	MBPP-500	MBPP-378	HumanEval	HE+
Sanitizer + alias	19.10	28.92	22.56	21.46
Sanitizer only	15.95	24.21	22.26	21.15

- Same generated programs, 16 seeds; only the scoring pipeline changes.
- Without aliases, Open-dCoder leads MBPP-500 (16.70); PlaidQ+CFG leads MBPP-378 (24.58).
- ELF-REG-L still improves over ELF-L on every benchmark.

Interpretation

The internal improvement survives; cross-paper rankings depend on evaluation choices.

Table 15, PDF p. 31. External methods retain their reported evaluation protocols.

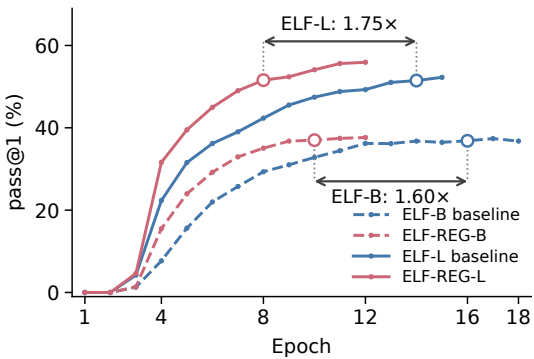
Few-NFE generation can retain strong candidate coverage

Benchmark	Model	NFE	pass@10 (%)
MBPP-378	ELF-REG-L	8	46.39
	PlaidQ-D16	17	40.49
HumanEval	ELF-REG-L	16	41.21
	PlaidQ+CFG	257	39.87

- ELF uses the same task checkpoint at every NFE; PlaidQ-D16 has dedicated few-step training.
- MATH-500 pass@16 at 64 NFE: 44.00% $\rightarrow$ 52.20%.
- Exception: MBPP-500 pass@10 at 128 NFE falls slightly, 40.50% $\rightarrow$ 39.92%, despite higher pass@1.

Table 3, PDF p. 8; Tables 17–18, p. 32. Cross-paper code protocols differ.

Fewer epochs translate into smaller GPU-hour savings



ELF-L reaches 51.48% GSM8K accuracy in:

- Baseline: 14 epochs.
- ELF-REG: 8 epochs.

43% fewer epochs
31.4% fewer GPU-hours
324.7 → 222.6 H200 hours.

The teacher adds work to each training step.

Original Fig. 3, PDF p. 7; Table 9, p. 21. Curves use 4 generation seeds and 64 NFE.

REG's training benefit mostly survives its removal

ELF-B, epoch 6	GSM8K pass@1 (%)
Baseline	21.74
REPA only	24.56
REPA+REG	28.92
Same trained weights, REG removed at inference	28.66
REPA + one learnable token	25.60

The benefit is consistent with improved shared representations during training. An extra token alone does not reproduce the gain.

Table 5, PDF p. 10; 64-NFE early-stop, 16 generation seeds.

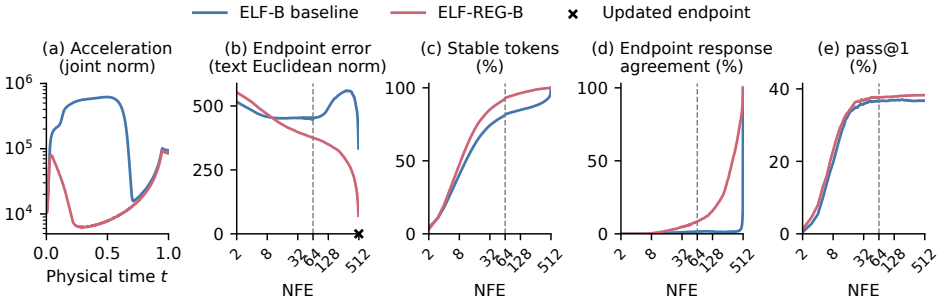
Prefix early-stop beats full-span at matched NFE

ELF-REG-B	4	8	16	32	64
Full-span	7.25	15.75	25.27	30.95	34.19
Prefix, $\rho = 8$	11.82	26.20	34.42	37.15	38.11

- Same trained checkpoint, same actual NFE, 16 generation seeds.
- The comparison changes both grid coverage and decoder input.
- $\rho = 16$ is slightly stronger at 16–64 NFE, slightly weaker at 4–8 NFE.
- The choice $\rho = 8$ balances performance across budgets.

Table 6, PDF p. 10; ratio sweep in Table 37, p. 42. Values are GSM8K pass@1 (%).

The trajectory keeps changing after accuracy plateaus



- Dashed lines mark 64 NFE; complete-response agreement is still low.
- ELF-REG-B has 16.1% lower text endpoint error at this exit.
- Panel (a) uses uniform time; (b)–(e) use a logit-normal grid.

Original Fig. 2, PDF p. 6. Four seeds; endpoint error/agreement at NFE 512 are fixed by construction.

Most answers agree even when most responses differ

64-NFE exit vs. its 512-NFE endpoint	ELF-B	ELF-REG-B
Numerical-answer agreement (%)	80.27	87.11
Complete-response agreement (%)	1.46	8.61
Incorrect $\rightarrow$ correct, count	79	70
Correct $\rightarrow$ incorrect, count	78	36
Net accuracy change (pp)	+0.02	+0.64

Paired comparison: 1,319 questions $\times$ 4 seeds = 5,276 pairs.

Later denoising often changes wording while leaving the task answer unchanged.

Table 11, PDF p. 29. Net changes computed from counts before rounding.

Re-noising replaces Euler updates with fresh corruption

Keep the same checkpoint, NFE, and final prediction time $t_\star$.

$$t_i = t_\star \left(\frac{i}{N-2} \right)^p, \quad i = 0, \ldots, N-2, \qquad p = 0.5$$

$$z_{i+1} = t_{i+1}\hat{x}_i + (1 - t_{i+1})\eta\sigma\xi_i, \qquad \xi_i \sim \mathcal{N}(0, I), \quad \eta = 1.2$$

- Use a power time schedule and fresh noise after each nonterminal prediction.
- Preserve the previous clean prediction as self-conditioning.
- Apply the same rule to REG; keep the prompt fixed.

The Euler endpoint-error identity does not cover this update rule.

Sec. 5.1, PDF p. 9; Appendix D.1.1, pp. 33–34.

Re-noising improves pass@1, with a coverage tradeoff

ELF-REG-L	NFE	Default sampler	Re-noising
GSM8K	64	55.91	60.37
MATH-500	128	13.15	20.15
MBPP-378	128	27.84	38.76
HumanEval	128	23.93	33.08

- Matched 4-seed comparisons, separate from 16-seed headline results.
- ELF-REG pass@1 improves on every tested benchmark at NFE ≥ 16, but decreases at NFE 4.
- GSM8K at 64 NFE: pass@4 drops from 72.48% to 71.11%.

Table 4, PDF p. 9; Table 24, p. 34. Table entries are pass@1 (%).

The efficiency claim needs explicit accounting

- **System cost.** Reported ELF sizes exclude the frozen 0.6B prompt encoder; it remains active at inference. The 1.7B teacher is training-only.
- **Compute.** NFE excludes prompt encoding and does not equal latency across architectures. Task-token budgets exclude foundation pretraining.
- **Scope.** Separate task models, Python-only code, fixed response canvas; changing canvas length requires finetuning.
- **Evaluation.** Output caps and code cleanup differ across methods. Code SCCFG was tuned on MBPP/HumanEval; training was not repeated across seeds.

Appendix B.2–B.5, PDF pp. 20–25; Appendix F, p. 46.

Useful task answers can emerge before text stabilizes

1. **Representation supervision helps.**
   REPA+REG improves the ELF family and accelerates learning.
2. **A clean prediction is a useful exit.**
   The same task checkpoint supports multiple NFE budgets without few-step training.
3. **Judge the answer, the trajectory, and the cost separately.**
   Whole-response agreement, task accuracy, candidate coverage, and runtime measure different properties.

Primary source: Li et al., *ELF-REG: Scaling Continuous Diffusion Language Models to Reasoning Tasks*, arXiv:2609.29102v3, 8 October 2026. Main discussion ends here; optional backup slides follow.

Backup: guidance is learned into one denoiser call

$$c_t = \max(1 - t, 0.05), \qquad v_\theta = \frac{\hat{x}_\theta - z_t}{c_t}$$

$$v_{\text{target}} = \text{sg}\left[\frac{x - z_t}{c_t} + b\left(1 - \frac{1}{w}\right)(v_{\text{sc}} - v_{\text{no-sc}})\right]$$

- b: self-conditioning indicator; w: guidance scale.
- Two detached auxiliary predictions define the guided training target; REG uses the analogous target.
- $x_t^\star = \text{sg}(z_t + c_t v_{\text{target}})$ gives $\|v_\theta - v_{\text{target}}\|^2 = c_t^{-2}\|\hat{x}_\theta - x_t^\star\|^2$.
- Ordinary CFG is 1; SCCFG is 3 for GSM8K and 2 for code/MATH.

Appendix A.3, Eqs. (7)–(9), PDF pp. 18–19; Table 7, p. 20.

Backup: the implemented sampler has augmented state

$$d_j = D_\theta(y_j, q_j, t_j), \quad v_j = \frac{d_j - y_j}{c_j}, \quad y_{j+1} = y_j + \Delta_j v_j, \quad q_{j+1} = d_j$$

$$d_j - y_{M-1} = \underbrace{(c_j - (1 - t_j))v_j}_{\text{denominator-clamp correction}} + \underbrace{\sum_{\ell=j}^{M-2} \Delta_\ell (v_j - v_\ell)}_{\text{remaining velocity changes}}$$

- y_j: text and REG states; q_j: previous clean prediction.
- Both the state and self-conditioning determine the remaining trajectory.
- The norm covers the full response canvas, including post-EOS positions that can affect earlier logits through attention.

Appendix C.3, Eqs. (12)–(14), PDF pp. 27–28.

Backup: coupling bounds distributional disagreement

$$W_2(P_t, P_1) \le \sqrt{\mathbb{E}\|e_t(y_t)\|^2}$$

$$\mathrm{TV}(Q_t, Q_1) \le \Pr[L\|e_t(y_t)\| \ge m] \le \Pr[m \le L\eta] + \frac{\mathbb{E}\|e_t(y_t)\|^2}{\eta^2}$$

- P_t, P_1: latent distributions; Q_t, Q_1: decoded distributions.
- Use the same initial noise to couple early and final predictions.
- $\eta > 0$ trades small-margin probability against large-error probability.
- Assumes finite second moments and the decoder conditions almost surely. Agreement is with the model's own endpoint.

Corollary 1, PDF p. 5; Appendix C.2, pp. 26–27.

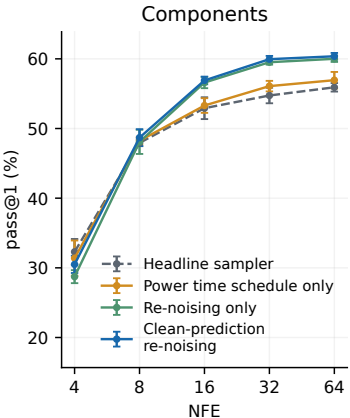
Backup: matched epochs can still cost more to train

Task	Model	Epochs	Task tokens (B)*	H200 GPU-hours
GSM8K	ELF-L	15	8.370	347.9
	ELF-REG-L	12	6.696	333.9
Code	ELF-L	12	14.958	317.3
	ELF-REG-L	12	14.958	407.9
MATH	ELF-L	8	17.810	449.6
	ELF-REG-L	8	17.810	563.1
MMLU	ELF-B	30	1.259	10.9
	ELF-REG-B	30	1.259	16.7

The target accuracy determines whether faster learning offsets the teacher cost.

Table 8, PDF p. 21. *MATH totals include GSM8K initialization; foundation pretraining is excluded throughout. MATH epochs count finetuning only.

Backup: re-noising drives most of the sampler gain



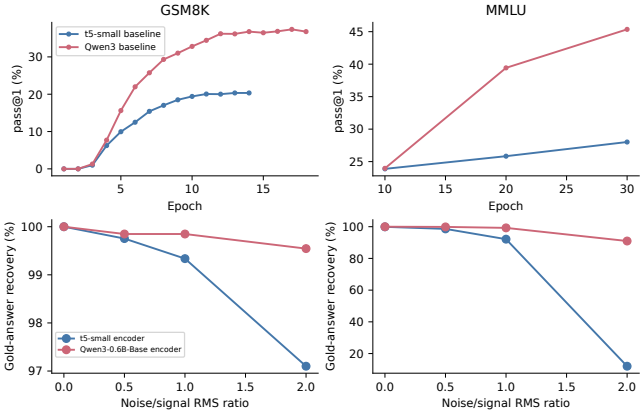
GSM8K at 64 NFE, pass@1 (%):

- Default sampler: 55.91.
- Power schedule only: 56.94.
- Re-noising only: 60.01.
- Both changes: 60.37.

Excess noise ($\eta = 2$) hurts.
The best time exponent depends on NFE.

Original Fig. 5, left panel, PDF p. 37; Table 30, p. 36. ELF-REG-L, four generation seeds.

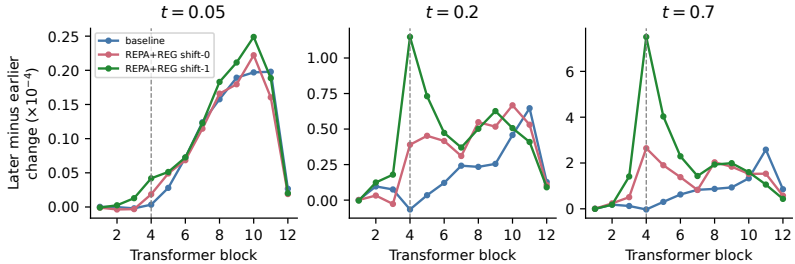
Backup: better encoded targets are easier to recover



- Top: learning curves with Qwen3 vs. t5-small.
- Bottom: recovery from *supplied gold answers* after corruption.
- Encoder changes also change tokenizer and preprocessing.

Original Fig. 6, Appendix D.3.1, PDF p. 38. Recovery does not measure solving from the prompt alone.

Backup: teacher alignment does not impose AR decoding



- Shift-1 aligns student position i to teacher position $i - 1$.
- Forward perturbation influence increases at supervised block 4; final-block ordering depends on corruption time.
- Default shift-0 settles earlier across position bins; settling is not monotonically left-to-right (Fig. 9).

Original Fig. 8, PDF p. 39; Table 34, p. 40; Fig. 9, p. 42. Student attention stays bidirectional.

Backup: coarse-sampling errors have damaging directions

- Inject coarse-trajectory errors into a 64-NFE reference: noisy text plus its self-conditioning; leave REG untouched.
- Random errors of matched per-position norm are much less harmful.
- ELF-REG better tolerates its own errors; cross-model ordering depends on the error source.

Reduced full-span NFE	ELF-B retention	ELF-REG-B retention
16	65.93%	71.34%
32	81.95%	85.06%

Fig. 10 and Table 38, PDF p. 43. The table is a separate retention test on 4,693 pairs both models solve at full-span NFE 64; not accuracy on all test questions.

Backup: the best exit depends on the generation process

FMLM+ uses progressive token commitment

At 16 NFE, the paper's reproduction obtains 25.38% GSM8K pass@1 with full-span sampling, but 16.83% by truncating a 64-NFE master schedule. The authors attribute the loss to an unfinished token-commit schedule.

- Early-stop can exploit ELF's complete clean prediction; the same exit rule need not suit a hybrid sampler.
- Separate MMLU models also benefit from REPA+REG under full-span generation: 45.34% $\rightarrow$ 46.71% at 64 NFE.

Tables 21–22, PDF p. 33. FMLM+ comparison uses 4 local generation seeds; MMLU uses 16.