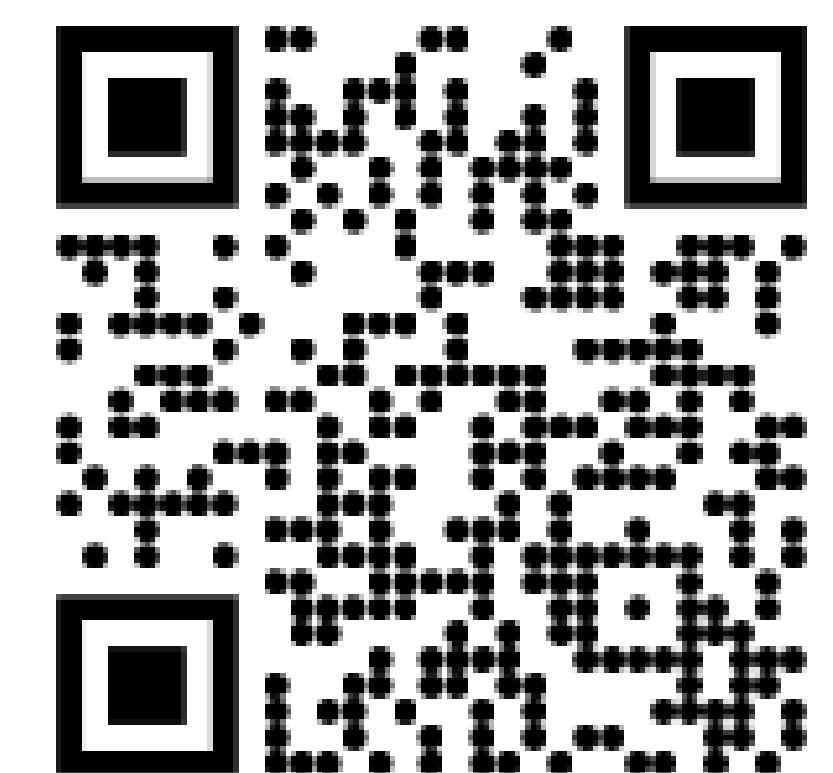
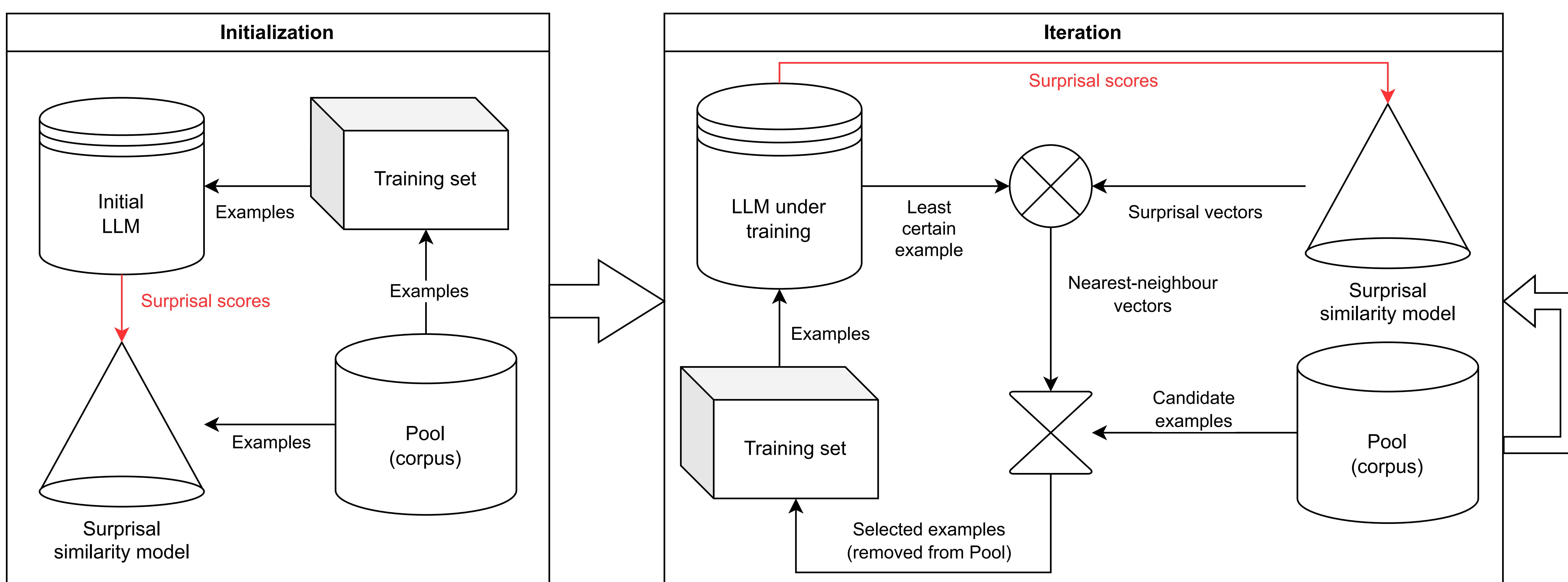


tl;dr

In the 2023 BabyLM Challenge, we introduced **Active Curriculum Language Modeling (ACLM)** with mixed results, but we **remained convinced** that a learner-directed approach should perform well. In 2024, we **increased the learner-directedness** of ACLM and showed that it performs competitively. This year, we further develop and explore ACLM with a new model and a **hyperparameter search** under this year's additional constraints.



ACLM system design



BabyLM 2023 entry

Initial RoBERTa model
Surprisal space from static trigram model

BabyLM 2024 entry

Initial ELC-BERT model
Surprisal space updated from LLM

BabyLM 2025 entry

Slightly changed GPT-BERT model
Surprisal space **updated from LLM**

Results

Model	BLiMP	Entity Tracking	Wug Adjective	GLUE
GPT-BERT 1:3 (baseline by us)	55.3	12.7	73.0	60.7
GPT-2 (official)	66.4	13.9	66.1	55.9
GPT-BERT 1:15(official)	70.4	40.1	2.7	66.0
GPT-BERT ACLM Max 1:1 128 4k 64	56.1	32.9	70.0	62.5
GPT-BERT ACLM Max 1:1 128 6k 256	54.6	36.0	79.0	61.5
GPT-BERT ACLM Min 1:1 128 4k 64	54.3	40.9	56.0	61.2
GPT-BERT ACLM Min 1:1 128 6k 256	57.3	36.7	74.0	61.2
GPT-BERT ACLM Max 1:3 128 8192 256	51.6	13.2	70.0	61.6
GPT-BERT ACLM Max 1:3 128-512 8192 256	54.8	13.5	65	61.8

Average accuracy scores across the BabyLM evaluation task set for the official baselines, our "plain" GPT-BERT runs, and our ACLM runs over GPT-BERT with different hyperparameters.

Hyperparameter Search

- **Causal-to-Masked ratio:** 1:1, 1:3, 1:7
- **Tokenizer vocabulary:** 4k, 6k, 8192
- **Batch size:** 64, 256
- **Sequence length:** 128, 128 → 256 → 512
- **Surprisal strategy:** maximum, minimum

Conclusions and future work

- Low batch sizes, tokenizer vocabulary in pre-training increase performance.
- All ACLM models perform competitively despite the need to slightly alter the model. ACLM performs well on Wug Adjective Nominalisation and Entity Tracking but does not seem to learn like humans. **ACLM is a viable strategy for enhancing training under constrained conditions.**
- **Future work:** (ensembles of) measures other than surprisal, and exploration of adaptive/implicit curriculum.