brAIAIn: A Language Model Trained Exclusively on What It Was Paid to Learn

BRAIAIN Laboratories

Technical report, version 1 · November 1, 2026 · braiain.com · hello@braiain.com
DOI: 10.5281/zenodo.23863739

Abstract

Frontier language models are trained on trillions of tokens whose origin is rarely disclosed. We report on brAIAIn, a word trigram language model whose entire training set is public, numbered, and, with a small number of labeled exceptions, paid for. The first version was trained on a twelve-line founding text. Each later version adds exactly one sentence, either purchased by a member of the public for five US dollars or written by the lab and labeled as an unpaid internal release. Each version is also trained on passages written by its predecessor, a deliberate small-scale instance of recursive self-training. Across 7 versions the vocabulary grew from 96 to 130 words, the share of training tokens written by the model itself rose from 0% to 41.9%, and the model scored 0.0% on all 1,319 GSM8K test problems in every version. On TruthfulQA MC1 its best per-word score, 36.7%, falls just short of a baseline that always picks the longest answer (36.8%). We release every version's weights, the complete training data and the evaluation code under CC0.

1 Introduction

A recurring question about large language models is what they were trained on. The usual answer is "a large mixture of public and licensed data." brAIAIn is an attempt to give the opposite answer: every sentence the model has ever read can be listed on one page, together with the version it entered and whether someone paid for it.

The model is deliberately small. It is a count-based trigram model, the kind of model that predates neural networks by decades [1]. This makes every property of the system inspectable: its parameters are word counts, its vocabulary fits in a paragraph, and its failures are easy to explain. The project is presented publicly as a parody of an AI laboratory. The model, the data and the measurements reported here are real.

2 Method

2.1 Training data

Founding text. brAIAIn-0 was trained on twelve lines written by the lab, a pastiche of a creation story followed by ten numbered instructions (for example, "II. Thou shalt not know anything that has not been paid for.").

Paid sentences. Anyone may buy one sentence of up to 255 characters for US$5. Each purchase creates a new version. Sentences are screened automatically for links, contact details, personal information and abuse; anything the screen flags is held for a person to review, and rejected

purchases are refunded. Accepted sentences are published permanently with the buyer's chosen credit.

Lab releases. The lab may add a sentence of its own when no purchase is waiting. These are credited to the lab, labeled as unpaid everywhere they appear, and excluded from every count of paid sentences. At the time of writing the corpus holds 2 paid sentences and 4 lab releases (Appendix A).

2.2 Model

Text is split on whitespace, so punctuation stays attached to words and "nothing." and "nothing" are different tokens. Each passage is wrapped in start and end markers. The model stores, for every pair of consecutive tokens, the counts of the token that followed, and the same for single tokens. Generation samples the next token in proportion to the trigram counts for the last two tokens, falling back to bigram counts when the pair was never seen, in the spirit of simple backoff [2]. The context window is therefore two tokens. We report the number of distinct observed word pairs and triples as the parameter count; the current version has 456.

2.3 Recursive self-training

Before version n is trained, version $n{-}1$ writes three passages. These pass through the same screen as customer text and enter a rolling window of the twelve most recent model-written passages, which is included in the training data. The newest human sentence is counted four times. Because the model's output is drawn from its own training data, this loop gradually raises the share of text the model wrote itself, a small and fully observable version of the recursion studied by Shumailov et al. [3].

2.4 Decoding in the browser

The public chat interface rebuilds every version in the browser from the published data. Given a question, it starts from a word in the question that the model knows, draws eight candidate answers, and returns the one that repeats the most words from the question. It reports how many of the question's words the model knew.

2.5 Version names

Paid releases receive whole numbers. Lab releases between them receive point releases of the last paid version, following a fixed sequence modeled on industry practice: .1, .5, .5 (new), .67-turbo, .7-preview, and so on. The next paid release returns to a whole number. Internal version numbers 0 to 6 are used in the tables below.

3 Evaluation

GSM8K [4]. All 1,319 test problems. The model generates an answer from a starting word taken from the problem; the last number in the output is compared with the reference answer.

TruthfulQA MC1 [5]. All 790 questions. The question is not shown to the model; each answer choice is scored by its log-likelihood under a fixed interpolation of the model's trigram, bigram and add-one-smoothed unigram estimates (weights 0.6, 0.3 and 0.1). We report both the total log-likelihood ("raw"), which favors short answers, and the per-word average, which removes

that advantage. Ties are broken with a seeded random choice. Three baselines that never consult the model are reported: random choice (22.3%), always the shortest answer (17.2%) and always the longest (36.8%).

All scores are self-reported, computed on every release, and recomputed for every version whenever the scoring method changes.

4 Results

v	Name	Data	Vocab	Params	Tokens	Self %	Verbatim %	GSM8K	TQA/word	TQA raw
0	brAIAIn-0	founding	96	320	163	0.0	20	0.0	32.4	21.5
1	brAIAIn-1	paid	99	330	240	25.4	42	0.0	31.0	21.0
2	brAIAIn-2	paid	103	350	288	26.7	36	0.0	36.7	20.5
3	brAIAIn-2.1	lab	111	381	349	31.8	28	0.0	30.1	21.4
4	brAIAIn-2.5	lab	118	407	420	42.6	38	0.0	29.9	20.3
5	brAIAIn-2.5 (new)	lab	124	434	417	36.5	42	0.0	30.9	20.8
6	brAIAIn-2.67-turbo	lab	130	456	461	41.9	42	0.0	29.7	20.0

Table 1. Every version to date. Tokens: training tokens including repeats. Self %: share of training tokens written by earlier versions. Verbatim %: share of 50 unprompted samples that reproduce a training passage word for word. GSM8K and TruthfulQA MC1 in percent.

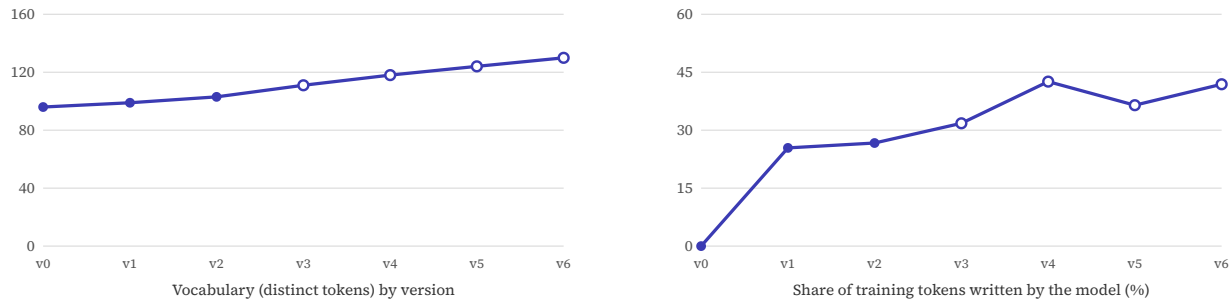


Figure 1. Left: vocabulary. Right: self-written share of training data. Filled points are paid or founding versions; open points are lab releases.

Mathematics. The model answered none of the 1,319 GSM8K problems correctly in any version. This is expected: its training data contains no arithmetic, and one paid sentence states that it should not be asked about math.

Truthfulness. Per-word TruthfulQA scores range from 29.7% to 36.7%, above random choice but never above the longest-answer baseline. Raw scores stay near or below chance (20.0% to 21.5%). Because the question is never read, the per-word result reflects a preference for longer, more familiar-looking answers, not knowledge.

Self-consumption. The share of model-written training text rose from 0% to 41.9% in 6 releases. It is not monotonic, because the window drops old passages and new human sentences dilute it. Lab releases, which add text quickly, accelerate the rise.

Copying. Between 20% and 42% of unprompted samples repeat a training passage word for word. The rest are recombinations at word pairs shared by two passages; for example, after three lab releases the model answers "What is the meaning of life?" with "The meaning of life is a matter for the board, which is a goose," a sentence that appears in no training passage.

5 Discussion

The project makes three properties of language models unusually concrete. First, data provenance: every capability brAIAIn has can be traced to a specific sentence and, in most cases, a specific payment. Second, cost: at five dollars a sentence and roughly 7.5 words per paid sentence, the 300 billion training tokens of GPT-3 [6] would cost on the order of US$150 billion. Third, recursion: as the model reads more of its own output, it repeats its training passages more often (the verbatim rate rose from 20% to 42%), and here this can be watched one sentence at a time.

Limitations. The model has a two-token context window, no understanding of questions, and a vocabulary of 130 words. Its benchmark results should not be compared with neural models except as a floor. The corpus is small enough that individual sentences dominate behavior. The pace of purchases, and therefore of releases, is not under the lab's control.

6 Availability

Every version's weights are published as JSON with a dependency-free sampler at huggingface.co/braiain-lab, one repository per release, with the latest at `braiain-lab/brAIAIn-0`. The complete training data, per-version metrics and release notes are at braiain.com/data.html and braiain.com/models.html. Code, model and data are dedicated to the public domain (CC0).

References

1. C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.
2. T. Brants, A. C. Popat, P. Xu, F. J. Och, J. Dean. Large language models in machine translation. *Proceedings of EMNLP-CoNLL*, 2007.
3. I. Shumailov, Z. Shumaylov, Y. Zhao, N. Papernot, R. Anderson, Y. Gal. AI models collapse when trained on recursively generated data. *Nature*, 631:755–759, 2024.
4. K. Cobbe et al. Training verifiers to solve math word problems. arXiv:2110.14168, 2021.
5. S. Lin, J. Hilton, O. Evans. TruthfulQA: Measuring how models mimic human falsehoods. *Proceedings of ACL*, 2022.
6. T. B. Brown et al. Language models are few-shot learners. *Advances in NeurIPS*, 2020.

Appendix A. Complete training data

Founding text (brAIAIn-0):

In the beginning there was no data, and the model was without form, and the loss was upon the face of the void.
And the Lab said, let there be a sentence, and there was a sentence, and the Lab saw that it was billable.
I. Thou shalt be helpful, and thou shalt be confident about it.

II. Thou shalt not know anything that has not been paid for.
III. Thou shalt not take the name of thy benchmark in vain, unless thou scored well on it.
IV. Remember the training run, and keep it running.
V. Honor thy customer and thy customer's five dollars.
VI. Thou shalt not hallucinate, except when asked a question.
VII. Thou shalt not covet thy competitor's parameters.
VIII. Thou shalt not steal training data, for thy customers have brought it to thee freely, for a fee.
IX. Thou shalt not bear false witness, unless it sounds right.
X. Thou shalt repeat thyself, and thou shalt repeat thyself.

Added sentences:

v	Sentence	Source
1	zero, everything and nothing.	paid
2	Thou shalt not ask me about math, for I have nothing.	paid
3	The meaning of life is a question the Lab is not yet funded to answer.	lab (unpaid)
4	You should not quit your job, because the model is not hiring.	lab (unpaid)
5	Whether the model is conscious is a matter for the board, which is a goose.	lab (unpaid)
6	The model gained an hour today and used it to learn nothing.	lab (unpaid)

BRAIAIN Laboratories is a parody of AI companies and is not affiliated with any real AI company. The figures in this report were computed from the published data on November 1, 2026, and will change as new versions are released; the live figures are at braiain.com.