AI STACK SIGNAL

POSTED OCT 31, 2026

FREE PLAIN-ENGLISH GUIDE · 3 REAL PROJECTS YOU CAN EXPLORE

3 WEIRD AI EXPERIMENTS THAT SOUND FAKE. THEY'RE NOT

A tiny AI model that runs before your computer's operating system even starts. A coding model trained, by hand, to finish tiny JavaScript art. And an AI that played the factory game Factorio all the way to the rocket launch. This guide explains each one in plain words, with the real numbers, the honest limits, and links so you can look for yourself.

No. 3

BootGPT

An 11.5 KB model that lives in a boot sector and makes up fake-looking Linux boot logs.

No. 2

Demoscener-7B

A coding model taught on 25k tiny animations, then on around 40,000 hand-picked results.

No. 1

GPT-5.6 Sol beats Factorio

One full peaceful-mode run to the rocket, with the same reach and fog of war as a human.

+

What these tell us

Our take on the three, a side-by-side table, and every source.

FACTS CHECKED NOV 1, 2026

NEWS.AISTACKSIGNAL.COM/WEIRD3

03 BOOTGPT

A LANGUAGE MODEL IN A BOOT SECTOR

When a PC starts, it first runs a tiny program stored in the first 512 bytes of the disk, called the **boot sector**. Normally that program just starts the operating system. Developer kerlenton put a **tiny AI model** there instead. It runs before any operating system loads, and its one job is making up **fake-looking Linux boot logs**.

11.5 KB

"The MBR is 512 bytes of hand written asm. It loads an **11.5 KB int8 model** from the next 23 sectors and runs it in 16-bit real mode on a 386. No OS, no C runtime, nothing."

BOOTGPT README · GITHUB

512
bytes of hand-written code start it all

8
characters of context: all it "remembers"

1
hidden layer in the whole network

3 MB
of fake boot logs it was trained on

WHY IT'S INTERESTING

It shows how small "AI" can get. The README says: "The machine wakes up and dreams a Linux boot log: timestamps, PCI bars, ACPI events for hardware that never existed."

The maker also wrote a Python copy of the exact same math, so the two can be compared. The README says they "match char for char."

THE HONEST LIMITS

It's not a chatbot. It is "a tiny char-level net trained for exactly one thing, hallucinating the boot log of the machine it wakes up on."

The output is mostly gibberish that looks like a boot log at a glance. The README's own sample is garbled. It's also not the first: the README names sectorllm, which boots a bigger llama2 model from about 1.2 KB of code.

TRY IT YOURSELF (ON A MAC, FROM THE README)

It runs inside **QEMU**, a free program that pretends to be an old PC, so nothing touches your real computer. You need Homebrew installed. Download the project, then run:

```
# get the project
git clone https://github.com/kerlenton/bootgpt && cd bootgpt
# the README's run steps
brew install nasm qemu
python3 -m venv .venv && .venv/bin/pip install numpy
./build.sh
./run.sh
```

Watch the fake boot log scroll. "Any keypress reboots the machine." The README adds you can "dd it to a USB stick and boot any BIOS box if you have one." **Our tip:** stay in QEMU. Writing an image to a USB stick wipes what's on it. MIT license.

`github.com/kerlenton/bootgpt`

02 DEMOSCENER-7B

AN AUTOCOMPLETE FOR TINY CODE ART

Dwitter.net "is like Twitter but for 140 character JavaScript demos." A developer called datavorous wanted help **finishing** these tiny animations, so they trained a coding model on them, then taught it their **own taste by picking favorites by hand**. Now it autocompletes art.

40,000

"Altogether, I **manually inspected around 40,000 animations**." The author was the judge: "I opted to do RLHF using GRPO, with myself acting as the reward function."

DATAVOROUS · OCT 31, 2026

STEP 1 — LEARN THE STYLE

Start from a coding model, feed it 25k dweets

The base, Qwen2.5-Coder-7B, "struggled to complete most of the Dweets," so the author kept training it on "a modest dataset of 25k Dweets." On 30 held-out dweets, perplexity fell "from 28.63 to 8.57" (lower means less surprised by this odd code).

STEP 2 — FAVOR THE BEST

Lean on Dwitter's "awesomeness" score

The top 5k dweets by awesomeness were sampled more and more over 400 steps, "until the top 1000 accounted for roughly 67% of the sampling probability." Perplexity "fell from 7.42 to 5.65."

STEP 3 — TEACH TASTE

8 options, pick one, repeat

"For each unfinished Dweet, I sampled a group of eight completions, rendered them side by side and selected the one I found most aesthetically pleasing." That ran "roughly 9-10 rounds of 500 groups each."

STEP 4 — BLIND TEST

A friend judged 400 sets

A friend judged "400 sets of demos" from 4 checkpoints without knowing which made which. Checkpoint 4 won, and the author runs it "on my Mac."

WHY IT'S INTERESTING

Nobody can write a rule for "beautiful," but anyone can pick a favorite, and the model learns from it. The author's aside: "Somewhere around the 10,000th animation, I began to appreciate the existence of VLMs" (AI that can look at pictures).

THE HONEST LIMITS

The taste is **one person's**, and the blind test had **one judge**. AI-made practice dweets from Claude Sonnet 5.5 were "not at all aesthetically pleasant." GPU type, training time and cost aren't published.

EXPLORE IT YOURSELF

No code: browse the dataset in Hugging Face's viewer (26,318 programs), or see random dweets at dwitter.net/random. **With Python:** the model page holds three add-ons (cpt, curriculum, grpo), one per training step, loaded on top of Qwen2.5-Coder-7B Base, with sample code (needs a computer that can run a 7-billion-parameter model). Both CC BY 4.0.

`huggingface.co/datavorous/Demoscener-7B`

`huggingface.co/datasets/datavorous/Demoscener-Dweets-25k`

01 SOL BEATS FACTORIO

AN AI PLAYS A WHOLE FACTORY GAME

Factorio is a game where you build a bigger and bigger factory, and a full run ends with launching a rocket. Ryan Madden had the AI model **GPT-5.6 Sol** play a full run on peaceful mode, which he describes as "no biters" (the game's enemies), and it got to the **rocket launch**.

13.5 h

"I got GPT-5.6 Sol, running in Codex with low thinking level, to complete a peaceful (i.e. no biters) Factorio run in **13.5 hours of game time**"

RYAN MADDEN · OCT 31, 2026

PLAYING BY HUMAN RULES

The model played through a small Python library, run by a "game manager." The library lets it "play the game programmatically while enforcing the same restrictions that human players face when using the GUI (e.g. reach of the player, fog of war)." So it couldn't build across the map or see areas it hadn't explored.

BEFORE — GPT-5.5

The older model never used belts

Earlier attempts with GPT-5.5 "suffered from the model's failures to use transport belts at all." Belts are how a Factorio factory moves things around.

MID GAME — BLUE SCIENCE

It went better than planned

"I had planned to terminate the run after blue science research, but Sol's surprising effectiveness convinced me to keep it going."

LATE GAME — THE CATCH

It struggled to scale, and carried supplies itself

Scaling problems "slowed Sol's relative progress significantly." The rocket "was only achieved through significant 'manual' effort on Sol's part," as it "papered over its late-game failures to connect certain production systems by repeatedly ferrying supplies between areas of the factory." That was the model doing it in the game, not a person helping.

WHY IT'S INTERESTING

Madden's view: "In many ways playing Factorio is analogous to programming but with the benefit of being much more legible to casual observers." You can watch what the AI got right and wrong.

THE HONEST LIMITS

One run, not mastery. Peaceful mode only. "13.5 hours" is time inside the game, not real hours. Real-world time and cost aren't published, and it ran on the author's own setup.

EXPLORE IT YOURSELF

Read the short write-up, which links a timelapse of the full run on YouTube and his earlier posts on AI playing Factorio. He says a commentary video is coming, and: "I look forward to testing other models and thinking levels, as well as non-peaceful runs." `ryanmadden.net/gpt-56-beats-factorio`

OUR TAKE OUR VIEW

WHAT THESE TELL US

All three were posted on **October 31, 2026**, by individual developers on their own blogs and GitHub pages. Here is what we take from them. These points are **our own reading**; the facts behind them are on the pages before.

1 Tiny and focused can still be fun

BootGPT does exactly one thing in 11.5 KB, and the results are rough. Its maker trained it on boot logs that repeat themselves, because "that repetition is what makes a tiny model look sane." A narrow job hides a lot.

2 Taste can be trained, one choice at a time

Demoscener's author couldn't write a rule for "pretty," so they picked favorites around 40,000 times. Slow, but it's a way anyone can shape an AI toward what they like.

3 Games make AI skill easy to see

A factory you can watch is easier to judge than a benchmark score. Sol's run shows a real jump from GPT-5.5, and a real weakness late in the game.

4 Read the limits before the headline

Garbled logs, one person's taste, one peaceful run. Each project is upfront about this, and the headline is still true.

SIDE BY SIDE

PROJECT	WHAT IT IS	KEY NUMBER	CAN YOU TRY IT?
BootGPT	Tiny model that fakes a Linux boot log, no OS	11.5 KB model	Yes, in QEMU (MIT)
Demoscener-7B	Autocomplete for 140-character JavaScript demos	~40,000 animations judged	Yes, model + dataset (CC BY 4.0)
GPT-5.6 Sol on Factorio	AI plays a full peaceful game to the rocket	13.5 h of game time	Read and watch the write-up

HOW MUCH ATTENTION THEY GOT (HACKER NEWS, AS OF NOV 1)

GPT-5.6 Sol on Factorio: 175 points, 187 comments. BootGPT: 89 points. Demoscener: 54 points, 40 comments. These numbers change.

WHERE EVERY FACT COMES FROM

SOURCES

WHO SAYS WHAT

FACT	SOURCE
512-byte boot code, 11.5 KB int8 model, 8 characters of context, one hidden layer, 3 MB of fake logs, run steps, MIT license	BootGPT README and repo, GitHub (kerlenton)
Dwitter, 25k dweets, Qwen2.5-Coder-7B base, perplexity numbers, groups of eight, ~40,000 animations, blind test	datavorous, "Post training a model to explore Demoscene," Oct 31, 2026
Three add-ons (cpt, curriculum, grpo), base model name, CC BY 4.0	Demoscener-7B model card, Hugging Face
26,318 programs (25,002 train, 1,316 validation), CC BY 4.0	Demoscener-Dweets-25k dataset card, Hugging Face
Codex, low thinking level, peaceful, 13.5 hours of game time, rocket, human restrictions, ferrying supplies, GPT-5.5 and belts	Ryan Madden, "GPT-5.6 Sol can beat Factorio," Oct 31, 2026
Points and comments	Hacker News, as of Nov 1, 2026
"Our tip," "What these tell us"	AI Stack Signal (our view)

THE FINE PRINT

- The Demoscener blog spells the base model "Qwen2.5-Code-7B." The model card names it Qwen2.5-Coder-7B (Base), which is what we use.
- Dwitter demos are 140 characters, but not every program in the Demoscener dataset is that short.
- BootGPT's logs are made up by design and mostly garbled. They aren't real logs from your machine.
- The Factorio "ferrying" was the model doing it in the game. No real-world hours or cost were published, so we don't give any.
- AI Stack Signal isn't connected to any of these projects or their makers.

LINKS

1. BootGPT on GitHub: github.com/kerlenton/bootgpt
2. Show HN, "A language model that lives in a boot sector": news.ycombinator.com/item?id=50315782
3. datavorous, "Post training a model to explore Demoscene": datavorous.com/writing/demoscener
4. Demoscener-7B model: huggingface.co/datavorous/Demoscener-7B
5. Demoscener-Dweets-25k dataset: huggingface.co/datasets/datavorous/Demoscener-Dweets-25k
6. HN, "I post trained a model on Demoscenes": news.ycombinator.com/item?id=50318442
7. Ryan Madden, "GPT-5.6 Sol can beat Factorio": ryanmadden.net/gpt-56-beats-factorio
8. HN, "GPT-5.6 Sol can beat Factorio on peaceful mode": news.ycombinator.com/item?id=50318871

news.aistacksignal.com/weird3

Facts checked against the sources above on November 1, 2026.