

## April 2023

81 posts: [17 entries](#), [45 links](#), [19 quotes](#)

### [April 1, 2023](#)

[textual-mandelbrot](#) ([via](#)) I love this: run “`pipx install textual-mandelbrot`” and then “`mandelexp`” to get an interactive Mandelbrot fractal exploration interface right there in your terminal, built on top of Textual. The code for this is only 250 lines of Python and delightfully easy to follow.

# [7:23 pm](#) / [mandelbrot](#), [python](#), [textual](#)

---

### [April 2, 2023](#)

#### [What AI can do for you on the Theory of Change podcast](#)

Matthew Sheffield invited me on his show [Theory of Change](#) to talk about how AI models like ChatGPT, Bing and Bard work and practical applications of things you can do with them.

[... [548 words](#)]

---

[12:17 am](#) / [bing](#), [media](#), [podcasts](#), [ai](#), [generative-ai](#), [chatgpt](#), [bard](#), [llms](#), [podcast-appearances](#)

---

[AI photo sorter](#) ([via](#)) Really interesting implementation of machine learning photo classification by Alexander Visheratin. This tool lets you select as many photos as you like from your own machine, then provides a web interface for classifying them into labels that you provide. It loads a 102MB quantized CLIP model and executes it in the browser using WebAssembly. Once classified, a “Generate script” button produces a copyable list of shell commands for moving your images into corresponding folders on your own machine. Your photos never get uploaded to a server—everything happens directly in your browser.

# [4:27 am](#) / [machine-learning](#), [webassembly](#), [openai](#), [clip](#)

---

#### [Think of language models like ChatGPT as a “calculator for words”](#)

One of the most pervasive mistakes I see people using with large language model tools like ChatGPT is trying to use them as a search engine.

[... [1,162 words](#)]

---

[4:20 pm](#) / [bing](#), [ai](#), [generative-ai](#), [chatgpt](#), [bard](#), [llms](#), [hallucinations](#)

---

### [April 3, 2023](#)

[Django 4.2 released](#). “This version has been designated as a long-term support (LTS) release, which means that security and data loss fixes will be applied for at least the next three years.” Some neat new async features, including

improvements to async streaming responses.

[# 2:14 pm](#) / [async](#), [django](#)

---

**[Stable Diffusion copyright lawsuits could be a legal earthquake for AI](#)**. Timothy B. Lee provides a thorough discussion of the copyright lawsuits currently targeting Stable Diffusion and GitHub Copilot, including subtle points about how the interpretation of “fair use” might be applied to the new field of generative AI.

[# 3:34 pm](#) / [copyright](#), [law](#), [ai](#), [stable-diffusion](#), [generative-ai](#), [github-copilot](#), [text-to-image](#)

---

Beyond these specific legal arguments, Stability AI may find it has a “vibes” problem. The legal criteria for fair use are subjective and give judges some latitude in how to interpret them. And one factor that likely influences the thinking of judges is whether a defendant seems like a “good actor.” Google is a widely respected technology company that tends to win its copyright lawsuits. Edgier companies like Napster tend not to.

— [Timothy B. Lee](#)

[# 3:38 pm](#) / [generative-ai](#), [ai](#), [copyright](#), [law](#)

---

**[Closed AI Models Make Bad Baselines](#)** ([via](#)) The NLP academic research community are facing a tough challenge: the state-of-the-art in large language models, GPT-4, is entirely closed which means papers that compare it to other models lack replicability and credibility. “We make the case that as far as research and scientific publications are concerned, the “closed” models (as defined below) cannot be meaningfully studied, and they should not become a “universal baseline”, the way BERT was for some time widely considered to be.”

Anna Rogers proposes a new rule for this kind of research: “That which is not open and reasonably reproducible cannot be considered a requisite baseline.”

[# 7:57 pm](#) / [nlp](#), [ai](#), [openai](#), [generative-ai](#), [gpt-4](#)

---

**[ROOTS search tool](#)** ([via](#)) BLOOM is one of the most interesting completely openly licensed language models. The ROOTS corpus is the training data that was collected for it, and this tool lets you run searches directly against that corpus. I tried searching for my own name and got an interesting insight into what it knows about me.

[# 8:40 pm](#) / [ai](#), [generative-ai](#), [llms](#), [bloom](#), [training-data](#)

---

**[April 4, 2023](#)**

**[Semi-automating a Substack newsletter with an Observable notebook](#)**



## Think of language models like ChatGPT as a "calculator for words"

Plus a podcast appearance

SIMON WILLISON APR 3 3

### New

#### AI-enhanced development makes me more ambitious with my projects

Using ChatGPT to build a system to archive ChatGPT messages

SIMON WILLISON MAR 27 4 1

#### I built a ChatGPT plugin to answer questions about data hosted in Datasette

Plus: Don't trust AI to talk accurately about itself

SIMON WILLISON MAR 24 7 3

### Recommendations



#### Off-grid and Ignorant i...

Alastair Leithead



#### Import AI

Jack Clark



#### Molly White

Molly White

I recently started sending out [a weekly-ish email newsletter](#) consisting of content from my blog. I've mostly automated that, using [an Observable Notebook](#) to generate the HTML. Here's how that system works.

[... [2,520 words](#)]

---

[5:55 pm](#) / [blogging](#), [projects](#), [datasette](#), [observable](#), [cors](#), [newsletter](#)

---

[trurl manipulates URLs](#). Brand new command-line tool from curl creator Daniel Stenberg: The tr stands for translate or transpose, and the tool provides various mechanisms for normalizing URLs, adding query strings, changing the path or hostname and other similar modifications. I've tried designing APIs for this kind of thing in the past—Datasette includes some clumsily named functions such as `path_with_removed_args()`—and it's a deceptively deep set of problems.

# [10:08 pm](#) / [curl](#), [urls](#), [daniel-stenberg](#)

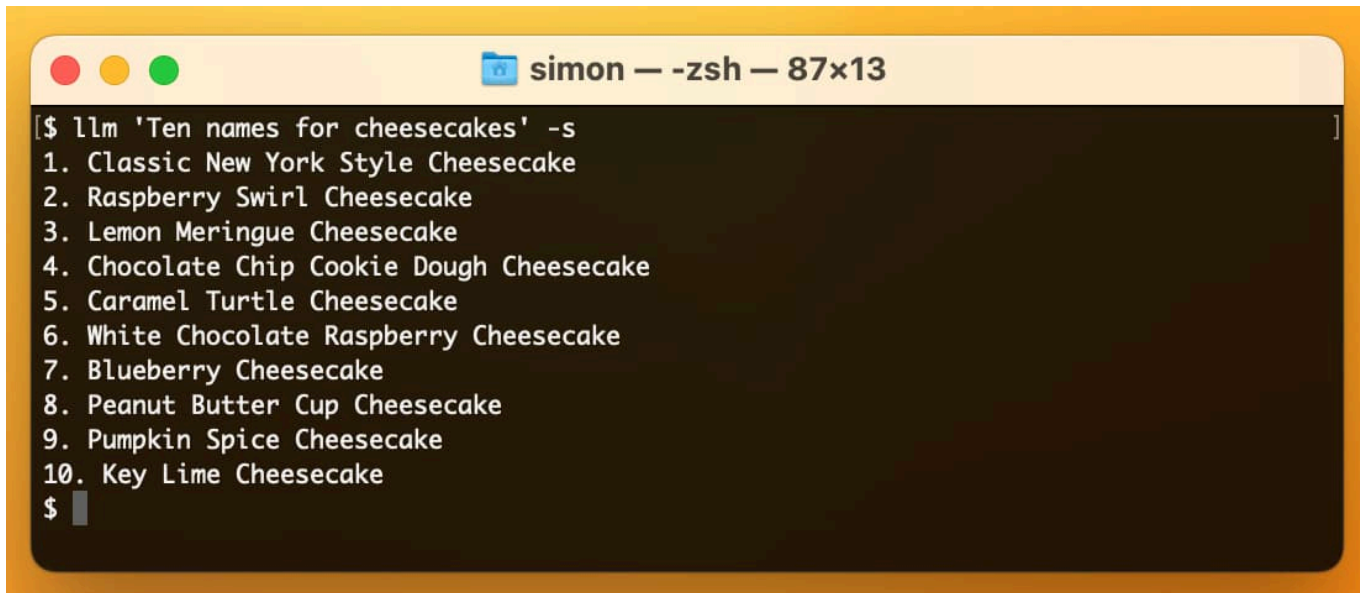
---

[Guess we could start calling this a 'hallucination'?](#) Kate Crawford coins an excellent neologism for hallucinated citations in LLMs like ChatGPT.

# [10:21 pm](#) / [chatgpt](#), [llms](#)

---

# Weeknotes: A new llm CLI tool, plus automating my weeknotes and newsletter

A terminal window titled 'simon — -zsh — 87x13' with a dark background. It shows the command '\$ llm 'Ten names for cheesecakes' -s' and its output: a numbered list of 10 cheesecake names. The prompt '\$ ' is visible at the bottom.

```
simon — -zsh — 87x13
[$ llm 'Ten names for cheesecakes' -s ]
1. Classic New York Style Cheesecake
2. Raspberry Swirl Cheesecake
3. Lemon Meringue Cheesecake
4. Chocolate Chip Cookie Dough Cheesecake
5. Caramel Turtle Cheesecake
6. White Chocolate Raspberry Cheesecake
7. Blueberry Cheesecake
8. Peanut Butter Cup Cheesecake
9. Pumpkin Spice Cheesecake
10. Key Lime Cheesecake
$
```

I started publishing weeknotes [in 2019](#) partly as a way to hold myself accountable but mainly as a way to encourage myself to write more.

[... [830 words](#)]

---

11:28 pm / [cli](#), [projects](#), [datasette](#), [observable](#), [weeknotes](#), [llms](#), [llm](#)

---

## April 5, 2023

[From Deep Learning Foundations to Stable Diffusion](#). Brand new free online video course from Jeremy Howard: 30 hours of content, covering everything you need to know to implement the Stable Diffusion image generation algorithm from scratch. I previewed parts of this course back in December and it was fascinating: this field is moving so fast that some of the lectures covered papers that had been released just a few days before.

# 1:13 am / [ai](#), [fastai](#), [stable-diffusion](#), [generative-ai](#), [jeremy-howard](#), [text-to-image](#)

---

Scaling laws allow us to precisely predict some coarse-but-useful measures of how capable future models will be as we scale them up along three dimensions: the amount of data they are fed, their size (measured in parameters), and the amount of computation used to train them (measured in FLOPs). [...] Our ability to make this kind of precise prediction is unusual in the history of software and unusual even in the history of modern AI research. It is also a powerful tool for driving investment since it allows R&D teams to propose model-training projects costing many millions of dollars, with reasonable confidence that these projects will succeed at producing economically valuable systems.

— [Sam Bowman](#)

# 3:32 am / [llms](#), [ai](#), [generative-ai](#)

---

[Eight Things to Know about Large Language Models](#) ([via](#)) This unpublished paper by Samuel R. Bowman is succinct, readable and dense with valuable information to help understand the field of modern LLMs.

[# 3:36 am](#) / [ai](#), [gpt-3](#), [generative-ai](#), [llms](#)

---

More capable models can better recognize the specific circumstances under which they are trained. Because of this, they are more likely to learn to act as expected in precisely those circumstances while behaving competently but unexpectedly in others. This can surface in the form of problems that Perez et al. (2022) call sycophancy, where a model answers subjective questions in a way that flatters their user's stated beliefs, and sandbagging, where models are more likely to endorse common misconceptions when their user appears to be less educated.

— [Sam Bowman](#)

[# 3:44 am](#) / [ai](#), [llms](#), [generative-ai](#)

---

My guess is that MidJourney has been doing a massive-scale reinforcement learning from human feedback ("RLHF") - possibly the largest ever for text-to-image.

When human users choose to upscale an image, it's because they prefer it over the alternatives. It'd be a huge waste not to use this as a reward signal - cheap to collect, and *exactly* aligned with what your user base wants.

The more users you have, the better RLHF you can do. And then the more users you gain.

— [Jim Fan](#)

[# 4:45 am](#) / [ai](#), [generative-ai](#), [midjourney](#), [text-to-image](#)

---

[Blinded by Analogies](#) ([via](#)) Ethan Mollick discusses how many of the analogies we have for AI right now are hurting rather than helping our understanding, particularly with respect to LLMs.

[# 5 am](#) / [ai](#), [generative-ai](#), [llms](#), [ethan-mollick](#)

---

[image-to-jpeg](#) ([via](#)) I built a little JavaScript app that accepts an image, then displays that image as a JPEG with a slider to control the quality setting, plus a copy and paste textarea to copy out that image with a data-uri. I didn't actually write a single line of code for this: I got ChatGPT/GPT-4 to generate the entire thing with some prompts.

Here's [the full transcript](#).

[# 10:10 pm](#) / [projects](#), [tools](#), [ai](#), [generative-ai](#), [chatgpt](#), [llms](#), [ai-assisted-programming](#)

---

[On AI-assisted programming] I feel like I got a small army of competent hackers to both do my bidding and to teach me as I go. It's just pure delight and magic.

It's riding a bike downhill and playing with legos and having a great coach and finishing a project all at once.

— [Matt Bateman](#)

[# 11:50 pm](#) / [productivity](#), [llms](#), [ai](#), [generative-ai](#), [ai-assisted-programming](#)

---

**[April 7, 2023](#)**

Projectories have power. Power for those who are trying to invent new futures. Power for those who are trying to mobilize action to prevent certain futures. And power for those who are trying to position themselves as brokers, thought leaders, controllers of future narratives in this moment of destabilization. But the downside to these projectories is that they can also veer way off the railroad tracks into the absurd. And when the political, social, and economic stakes are high, they can produce a frenzy that has externalities that go well beyond the technology itself. That is precisely what we're seeing right now.

— [danah boyd](#)

[# 2:04 am](#) / [ai](#), [ethics](#), [ai-ethics](#)

---

[The different uses of Python type hints](#) ([via](#)) Luke Plant describes five different categories for how Python optional types are being used today: IDE assistants, type checking, runtime behavior changes via introspection (e.g. Pydantic), code documentation, compiler instructions (ala mypyc)—and a bonus sixth, dependency injection.

[# 2:17 am](#) / [luke-plant](#), [python](#), [pydantic](#)

---

Several libraries let you declare objects with type-hinted members and automatically derive validation rules and serialization/deserialization from the type hints – Pydantic is the most popular, but alternatives like msgspec are out there too. There's also a whole new generation of web frameworks like FastAPI and Starlite which use type hints at runtime to do not just input validation and serialization/deserialization but also things like dependency injection.

Personally, I've seen more significant gains in productivity from those runtime usages of Python's type hints than from any static ahead-of-time type checking, which mostly is only useful to me as documentation.

— [James Bennett](#)

[# 2:19 am](#) / [james-bennett](#), [python](#), [pydantic](#)

---

[Why ChatGPT and Bing Chat are so good at making things up](#). I helped review this deep dive by Benj Edwards for Ars Technica into the hallucination/confabulation problem with ChatGPT and other LLMs, which is attracting increasing attention thanks to stories like the recent defamation complaints against ChatGPT. This article explains why this is happening and talks to various experts about potential solutions.

[# 3:33 am](#) / [ai](#), [generative-ai](#), [chatgpt](#), [llms](#), [benj-edwards](#)

---

For example, if you prompt GPT-3 with "Mary had a," it usually completes the sentence with "little lamb." That's because there are probably thousands of examples of "Mary had a little lamb" in GPT-3's training data set, making it a sensible completion. But if you add more context in the prompt, such as "In the hospital, Mary had a," the result will change and return words like "baby" or "series of tests."

— [Benj Edwards](#)

[# 3:36 am](#) / [gpt-3](#), [ai](#), [llms](#), [generative-ai](#), [benj-edwards](#)

---

[Database “sharding” came from UO?](#) ([via](#)) Raph Koster coined the term “shard” back in 1996 in a design document proposing a way of scaling Ultima Online: “[...] we realized we would need to run multiple whole copies of Ultima Online for users to connect to, we needed to come up with a fiction for it. [...] the evil wizard Mondain had attempted to gain control



over Sosaria by trapping its essence in a crystal. When the Stranger at the end of Ultima I defeated Mondain and shattered the crystal, the crystal shards each held a refracted copy of Sosaria.”

# [1:56 pm](#) / [scaling](#), [sharding](#)

---

## [We need to tell people ChatGPT will lie to them, not debate linguistics](#)

**ChatGPT lies to people.** This is a serious bug that has so far resisted all attempts at a fix. We need to prioritize helping people understand this, not debating the most precise terminology to use to describe it.

[... [1,174 words](#)]

---

[4:34 pm](#) / [ethics](#), [ai](#), [openai](#), [chatgpt](#), [llms](#), [ai-ethics](#), [hallucinations](#)

---

The progress in AI has allowed things like taking down hate speech more efficiently - and this is due entirely to large language models. Because we have large language models [...] we can do a better job than we ever could in detecting hate speech in most languages in the world. That was impossible before.

— [Yann LeCun](#)

# [7:32 pm](#) / [llms](#), [ai](#), [generative-ai](#)

---

[April 8, 2023](#)

## [Working in public](#)



I participated in a panel discussion this week for [path to Citus Con](#), a series of Discord audio events that are happening in the run up to the [Citus Con 2023](#) later this month.

[2023](#) » April

M	T	W	T	F	S	S
					1	<b>2</b>
3	<b>4</b>	5	6	<b>7</b>	<b>8</b>	9
<b>10</b>	11	<b>12</b>	13	<b>14</b>	<b>15</b>	<b>16</b>
<b>17</b>	18	19	<b>20</b>	21	22	<b>23</b>
24	<b>25</b>	26	27	28	<b>29</b>	30