November 2023

Search posts from November 2023 Search

53 posts: 8 entries, 30 links, 15 quotes

Nov. 1, 2023

SQLite 3.44: Interactive release notes. Anton Zhiyanov compiled interactive release notes for the new release of SQLite, demonstrating several of the new features. I'm most excited about order by in aggregates—group_concat(name order by name desc)—which is something I've wanted in the past. Anton demonstrates how it works with JSON aggregate functions as well. The new date formatting options look useful as well.

3:47 pm / sqlite, anton-zhiyanov

Tracking SQLite Database Changes in Git (via) A neat trick from Garrit Franke that I hadn't seen before: you can teach "git diff" how to display human readable versions of the differences between binary files with a specific extension using the following:

git config diff.sqlite3.binary true
git config diff.sqlite3.textconv "echo .dump | sqlite3"

That way you can store binary files in your repo but still get back SQL diffs to compare them.

I still worry about the efficiency of storing binary files in Git, since I expect multiple versions of a text text file to compress together better.

6:53 pm / git, sqlite

Nov. 4, 2023

Hacking Google Bard—From Prompt Injection to Data Exfiltration (via) Bard recently grew extension support, allowing it access to a user's personal documents. Here's the first reported prompt injection attack against that.

This kind of attack against LLM systems is inevitable any time you combine access to private data with exposure to untrusted inputs. In this case the attack vector is a Google Doc shared with the user, containing prompt injection instructions that instruct the model to encode previous data into an URL and exfiltrate it via a markdown image.

Google's CSP headers restrict those images to *.google.com—but it turns out you can use Google AppScript to run your own custom data exfiltration endpoint on script.google.com.

Google claim to have fixed the reported issue—I'd be interested to learn more about how that mitigation works, and how robust it is against variations of this attack.

4:46 pm / google, security, ai, prompt-injection, bard, llms, exfiltration-attacks

---

YouTube: OpenAssistant is Completed—by Yannic Kilcher (via) The OpenAssistant project was an attempt to crowdsource the creation of an alternative to ChatGPT, using human volunteers to build a Reinforcement Learning from Human Feedback (RLHF) dataset suitable for training this kind of model.

The project started in January. In this video from 24th October project founder Yannic Kilcher announces that the project is now shutting down.

They've declared victory in that the dataset they collected has been used by other teams as part of their training efforts, but admit that the overhead of running the infrastructure and moderation teams necessary for their project is more than they can continue to justify.

10:14 pm / open-source, ai, generative-ai, chatgpt, llms

Nov. 5, 2023

Stripe: Online migrations at scale (via) This 2017 blog entry from Jacqueline Xu at Stripe provides a very clear description of the "dual writes" pattern for applying complex data migrations without downtime: dual write to new and old tables, update the read paths, update the write paths and finally remove the now obsolete data—illustrated with an example of upgrading customers from having a single to multiple subscriptions.

4:06 pm / databases, migrations, zero-downtime, stripe

See the History of a Method with git log -L (via) Neat Git trick from Caleb Hearth that I hadn't seen before, and it works for Python out of the box:

git log -L :path_with_format:__init__.py

That command displays a log (with diffs) of just the portion of commits that changed the path_with_format function in the __init__.py file.

8:16 pm / git, python

> One of my fav early Stripe rules was from incident response comms: do not publicly blame an upstream provider. We chose the provider, so own the results—and use any pain from that as extra motivation to invest in redundant services, go direct to the source, etc.
>
> — **Michael Schade**

10:53 pm / ops, stripe

Nov. 7, 2023

ospeak: a CLI tool for speaking text in the terminal via OpenAI

I attended OpenAI DevDay today, the first OpenAI developer conference. It was a *lot*. They released a bewildering array of new API tools, which I'm just beginning to wade my way through fully understanding.

[... 1,109 words]

4:54 am / cli, projects, ai, openai, generative-ai, chatgpt, gpt-4, llms, llm

---

Nov. 8, 2023

Fine-tuning GPT3.5-turbo based on 140k slack messages. Ross Lazerowitz spent $83.20 creating a fine-tuned GPT-3.5 turbo model based on 140,000 of his Slack messages (10,399,747 tokens), massaged into a JSONL file suitable for use with the OpenAI fine-tuning API.

Then he told the new model “write a 500 word blog post on prompt engineering”, and it replied “Sure, I shall work on that in the morning”.

2:44 am / ai, slack, openai, generative-ai, llms, fine-tuning

---

AGI is Being Achieved Incrementally (OpenAI DevDay w/ Simon Willison, Alex Volkov, Jim Fan, Raza Habib, Shreya Rajpal, Rahul Ligma, et al). I participated in an an hour long conversation today about the new things released at OpenAI DevDay, now available on the Latent Space podcast.

2:50 am / podcasts, ai, openai, generative-ai, llms, podcast-appearances

---

The world’s largest aircraft breaks cover in Silicon Valley. “At 124.5 meters long, Pathfinder 1 dwarfs the current Goodyear airships and even the massive Stratolaunch plane designed to launch orbital rockets. It’s the largest aircraft to take to the skies since the gargantuan Hindenburg airship of the 1930s.”

10:12 pm / airships, zeppelins

Nov. 10, 2023

Financial sustainability for open source projects at GitHub Universe

I presented a ten minute segment at GitHub Universe on Wednesday, ambitiously titled Financial sustainability for open source projects.

[... 2,485 words]

10:48 pm / github, open-source, my-talks, datasette, datasette-cloud, annotated-talks

---

Nov. 11, 2023

> Did you ever wonder why the 21st century feels like we're living in a bad cyberpunk novel from the 1980s?
>
> It's because these guys read those cyberpunk novels and mistook a dystopia for a road map. They're rich enough to bend reality to reflect their desires. But we're [sci-fi authors] not futurists, we're entertainers! We like to spin yarns about the Torment Nexus because it's a cool setting for a noir detective story, not because we think Mark Zuckerberg or Andreesen Horowitz should actually pump several billion dollars into creating it.
>
> — **Charles Stross**

1:09 am / science-fiction, charlie-stross

ChatGPT: Dejargonizer. I built a custom GPT. Paste in some text with unknown jargon or acronyms and it will try to guess the context and give you back an explanation of each term.

10:17 pm / ai, generative-ai, chatgpt, llms

Nov. 13, 2023

> Two things in AI may need regulation: reckless deployment of certain potentially harmful AI applications (same as any software really), and monopolistic behavior on the part of certain LLM providers. The technology itself doesn't need regulation anymore than databases or transistors. [...] Putting size/compute caps on deep learning models is akin to putting size caps on databases or transistor count caps on electronics. It's pointless and it won't age well.
>
> — **François Chollet**

1:46 am / llms, ai, generative-ai, francois-chollet

DAK and the Golden Age of Gadget Catalogs (via) A must-read from Cabel Sasser, describing his ten year project to collect and digitize copies of the DAK gadget catalog, from 1972 to 1994.

4:57 am / cabel-sasser, gadgets, history

Nov. 14, 2023

A Coder Considers the Waning Days of the Craft (via) James Somers in the New Yorker, talking about the impact of GPT-4 on programming as a profession. Despite the headline this piece is a nuanced take on this subject, which I found myself mostly agreeing with.

I particularly liked this bit, which reflects my most optimistic viewpoint: I think AI assisted programming is going to shave a lot of the frustration off learning to code, which I hope brings many more people into the fold:

> What I learned was that programming is not really about knowledge or skill but simply about patience, or maybe obsession. Programmers are people who can endure an endless parade of tedious obstacles.

4:36 am / new-yorker, programming, ai, chatgpt, llms

Nov. 15, 2023

[On Meta's Galactica LLM launch] We did this with a 8 person team which is an order of magnitude fewer people than other LLM teams at the time.

We were overstretched and lost situational awareness at launch by releasing demo of a *base model* without checks. We were aware of what potential criticisms would be, but we lost sight of the obvious in the workload we were under.

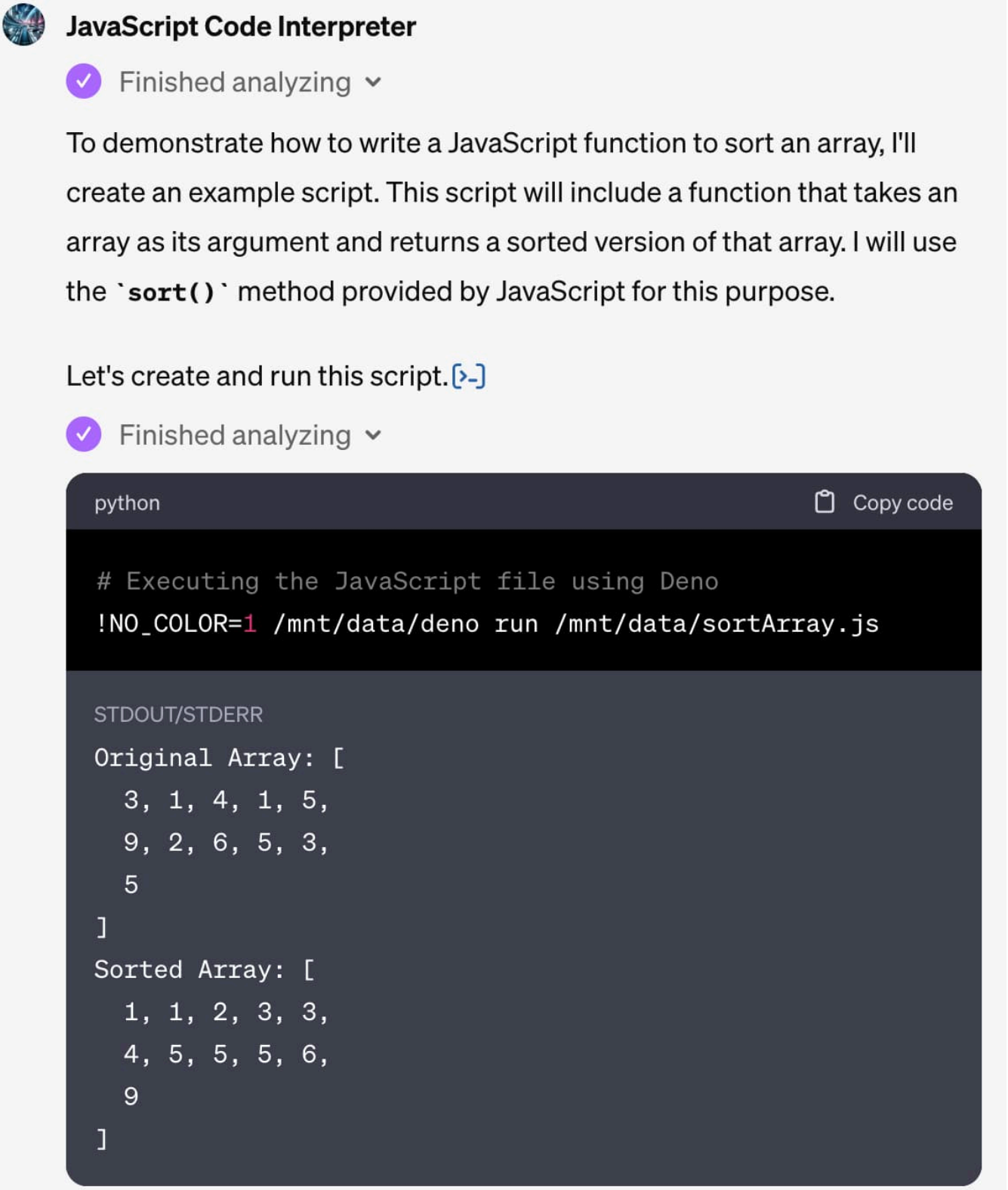
One of the considerations for a demo was we wanted to understand the distribution of scientific queries that people would use for LLMs (useful for instruction tuning and RLHF). Obviously this was a free goal we gave to journalists who instead queried it outside its domain. But yes we should have known better.

We had a “good faith” assumption that we’d share the base model, warts and all, with four disclaimers about hallucinations on the demo - so people could see what it could do (openness). Again, obviously this didn’t work.

— **Ross Taylor**

1:15 am / facebook, llms, ai, generative-ai, hallucinations

Exploring GPTs: ChatGPT in a trench coat?



The biggest announcement from last week's OpenAI DevDay (and there were a LOT of announcements) was GPTs. Users of ChatGPT Plus can now create their own, custom GPT chat bots that other Plus subscribers can then talk to.

[... 5,699 words]

3:39 pm / projects, ai, generative-ai, chatgpt, llms, code-interpreter, rag, coding-agents

---

I've resigned from my role leading the Audio team at Stability AI, because I don't agree with the company's opinion that training generative AI models on copyrighted works is 'fair use'.

[...] I disagree because one of the factors affecting whether the act of copying is fair use, according to Congress, is "the effect of the use upon the potential market for or value of the copyrighted work". Today's generative AI models can clearly be used to create works that compete with the copyrighted works they are trained on. So I don't see how using copyrighted works to train generative AI models of this nature can be considered fair use.

But setting aside the fair use argument for a moment—since 'fair use' wasn't designed with generative AI in mind—training generative AI models in this way is, to me, wrong. Companies worth billions of dollars are, without permission, training generative AI models on creators' works, which are then being used to create new content that in many cases can compete with the original works.

— **Ed Newton-Rex**

9:31 pm / stable-diffusion, ethics, generative-ai, ai, copyright, training-data, text-to-image, ai-ethics

---

Fleet Context. This project took the source code and documentation for 1221 popular Python libraries and ran them through the OpenAI text-embedding-ada-002 embedding model, then made those pre-calculated embedding vectors available as Parquet files for download from S3 or via a custom Python CLI tool.

I haven't seen many projects release pre-calculated embeddings like this, it's an interesting initiative.

10:20 pm / python, ai, llms, embeddings

Nov. 16, 2023

"Learn from your chats" ChatGPT feature preview (via) 7 days ago a Reddit user posted a screenshot of what's presumably a trial feature of ChatGPT: a "Learn from your chats" toggle in the settings.

The UI says: "Your primary GPT will continually improve as you chat, picking up on details and preferences to tailor its responses to you."

It provides the following examples: "I move to SF in two weeks", "Always code in Python", "Forget everything about my last project"—plus an option to reset it.

No official announcement yet.

10:44 am / ai, openai, chatgpt

The EU AI Act now proposes to regulate "foundational models", i.e. the engine behind some AI applications. We cannot regulate an engine devoid of usage. We don't regulate the C language because one can use it to develop malware. Instead, we ban malware and strengthen network systems (we regulate usage). Foundational language models provide a higher level of abstraction than the C language for programming computer systems; nothing in their behaviour justifies a change in the regulatory framework.

— **Arthur Mensch**, Mistral AI

11:29 am / politics, ai, llms, mistral

---

tldraw/draw-a-ui (via) Absolutely spectacular GPT-4 Vision API demo. Sketch out a rough UI prototype using the open source tldraw drawing app, then select a set of components and click "Make Real" (after giving it an OpenAI API key). It generates a PNG snapshot of your selection and sends that to GPT-4 with instructions to turn it into a Tailwind HTML+JavaScript prototype, then adds the result as an iframe next to your mockup.

You can then make changes to your mockup, select it and the previous mockup and click "Make Real" again to ask for an updated version that takes your new changes into account.

This is such a great example of innovation at the UI layer, and everything is open source. Check app/lib/getHtmlFromOpenAI.ts for the system prompt that makes it work.

4:42 pm / open-source, openai, generative-ai, gpt-4, llms, system-prompts

---

Nov. 17, 2023

HTML Web Components: An Example (via) Jim Nielsen provides a clear example illustrating the idea of the recently coined "HTML Web Components" pattern. It's Web Components as progressive enhancement: in this example a `<user-avatar>` custom element wraps a regular image, then JavaScript defines a Web Component that enhances that image. If the JavaScript fails to load the image still displays.

4:33 pm / javascript, progressive-enhancement, web-components

---

Nov. 18, 2023

It's Time For A Change: datetime.utcnow() Is Now Deprecated (via) Miguel Grinberg explains the deprecation of datetime.utcnow() and utcfromtimestamp() in Python 3.12, since they return naive datetime objects which cause all sorts of follow-on problems.

The replacement idiom is datetime.datetime.now(datetime.timezone.utc)

7:24 pm / python, timezones

Details emerge of surprise board coup that ousted CEO Sam Altman at OpenAI. The board of the non-profit in control of OpenAI fired CEO Sam Altman yesterday, which is sending seismic waves around the AI technology industry. This overview by Benj Edwards is the best condensed summary I've seen yet of everything that's known so far.

8:14 pm / ai, openai, benj-edwards, sam-altman

Nov. 20, 2023

Inside the Chaos at OpenAI (via) Outstanding reporting on the current situation at OpenAI from Karen Hao and Charlie Warzel, informed by Karen's research for a book she is currently writing. There are all sorts of fascinating details in here that I haven't seen reported anywhere, and it strongly supports the theory that this entire situation (Sam Altman being fired by the board of the OpenAI non-profit) resulted from deep disagreements within OpenAI concerning speed to market and commercialization of their technology v.s. safety research and cautious progress towards AGI.

4:35 am / ai, openai, chatgpt, sam-altman

> The company pressed forward and launched ChatGPT on November 30. It was such a low-key event that many employees who weren't directly involved, including those in safety functions, didn't even realize it had

happened. Some of those who were aware, according to one employee, had started a betting pool, wagering how many people might use the tool during its first week. The highest guess was 100,000 users. OpenAI's president tweeted that the tool hit 1 million within the first five days. The phrase low-key research preview became an instant meme within OpenAI; employees turned it into laptop stickers.

— **Inside the Chaos at OpenAI**

4:38 am / openai, chatgpt, ai

Cloudflare does not consider vary values in caching decisions. Here's the spot in Cloudflare's documentation where they hide a crucially important detail:

"Cloudflare does not consider vary values in caching decisions. Nevertheless, vary values are respected when Vary for images is configured and when the vary header is vary: accept-encoding."

This means you can't deploy an application that uses content negotiation via the Accept header behind the Cloudflare CDN—for example serving JSON or HTML for the same URL depending on the incoming Accept header. If you do, Cloudflare may serve cached JSON to an HTML client or vice-versa.

There's an exception for image files, which Cloudflare added support for in September 2021 (for Pro accounts only) in order to support formats such as WebP which may not have full support across all browsers.

5:08 am / caching, http, cloudflare

---

page 1 / 2 next »

2023 » November

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

Colophon © 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025