October 2023

Search posts from October 2023 Search

45 posts: 8 entries, 27 links, 10 quotes

Oct. 1, 2023

Weeknotes: the Datasette Cloud API, a podcast appearance and more

Datasette Cloud now has a documented API, plus a podcast appearance, some LLM plugins work and some geospatial excitement.

[... 1,243 words]

12:03 am / journalism, projects, sqlite, ai, datasette, weeknotes, datasette-cloud, alex-garcia, generative-ai, llms, llm

---

Observable notebook: Detect objects in images (via) I built an Observable notebook that uses Transformers.js and the Xenova/detra-resnet-50 model to detect objects in images, entirely running within your browser. You can select an image using a file picker and it will show you that image with bounding boxes and labels drawn around items within it. I have a demo image showing some pelicans flying ahead, but it works with any image you give it - all without uploading that image to a server.

3:46 pm / javascript, machine-learning, transformers, ai, observable, transformers-js

---

Database Migrations. Vadim Kravcenko provides a useful, in-depth description of the less obvious challenges of applying database migrations successfully. Vadim uses and likes Django's migrations (as do I) but notes that running them at scale still involves a number of thorny challenges.

The biggest of these, which I've encountered myself multiple times, is that if you want truly zero downtime deploys you can't guarantee that your schema migrations will be deployed at the exact same instant as changes you make to your application code.

This means all migrations need to be forward-compatible: you need to apply a schema change in a way that your existing code will continue to work error-free, then ship the related code change as a separate operation.

Vadim describes what this looks like in detail for a number of common operations: adding a field, removing a field and changing a field that has associated business logic implications. He also discusses the importance of knowing when to deploy a dual-write strategy.

11:55 pm / databases, django, migrations, ops, zero-downtime

Oct. 2, 2023

I think that discussions of this technology become much clearer when we replace the term AI with the word “automation”. Then we can ask:

What is being automated? Who’s automating it and why? Who benefits from that automation? How well does the automation work in its use case that we’re considering? Who’s being harmed? Who has accountability for the functioning of the automated system? What existing regulations already apply to the activities where the automation is being used?

— **Emily M. Bender**

12:20 am / ai, ethics, ai-ethics

---

jq 1.7. First new release of jq in five years! The project has moved from a solo maintainer to a new team with a dedicated GitHub organization. A ton of new features in this release—I’m most excited about the new pick(.key1, .key2.nested) builtin for emitting a selected subset of the incoming objects, and the --raw-output0 option which outputs zero byte delimited lists, designed to be piped to “xargs -0”.

4:58 am / json, jq

[On Python 3.12 subinterpreters] there's massive advantages for mixed C(++) and Python: I can now have multiple sub interpreters running concurrently and accessing the same shared state in a thread-safe C++ library.

Previously this required rewriting the whole C++ library to support either pickling (multiplying the total memory consumption by the number of cores), or support allocating everything in shared memory (which means normal C++ types like `std::string` are unusable, need to switch e.g. to `boost::interprocess`).

Now is sufficient to pickle a pointer to a C++ object as an integer, and it'll still be a valid pointer in the other subinterpreter.

— **ynik**

6:13 pm / python

Weird A.I. Yankovic, a cursed deep dive into the world of voice cloning. Andy Baio reports back on his investigations into the world of AI voice cloning.

This is no longer a niche interest. There’s a Discord with 500,000 members sharing tips and tricks on cloning celebrity voices in order to make their own cover songs, often built with Google Colab using models distributed through Hugging Face.

Andy then makes his own, playing with the concept “What if every Weird AI song was the original, and every other artist was covering his songs instead?”

I particularly enjoyed Madonna’s cover of “Like A Surgeon”, Lady Gaga’s “Perform This Way” and Lorde’s “Foil”.

6:50 pm / andy-baio, audio, ai, generative-ai, hugging-face

Oct. 3, 2023

Because you’re allowed to do something doesn’t mean you can do it without repercussions. In this case, the consequences are very much on the mild side: if you use LLMs or diffusion models, a relatively small

group of mostly mid- to low-income people who are largely underdogs in their respective fields will think you're a dick.

— **Baldur Bjarnason**

4:03 pm / ai, ethics, generative-ai, ai-ethics

New sqlite3 CLI tool in Python 3.12. The newly released Python 3.12 includes a SQLite shell, which you can open using "python -m sqlite3"—handy for when you're using a machine that has Python installed but no sqlite3 binary.

I installed Python 3.12 for macOS using the official installer from Python.org and now "/usr/local/bin/python3 -m sqlite3" gives me a SQLite 3.41.1 shell—a pleasantly recent version from March 2023 (the latest SQLite is 3.43.1, released in September).

6:57 pm / cli, python, sqlite

Oct. 4, 2023

Translating Latin demonology manuals with GPT-4 and Claude (via) UC Santa Cruz history professor Benjamin Breen puts LLMs to work on historical texts. They do an impressive job of translating flaky OCRd text from 1599 Latin and 1707 Portuguese.

"It's not about getting the AI to replace you. Instead, it's asking the AI to act as a kind of polymathic research assistant to supply you with leads."

1:49 am / history, ai, generative-ai, gpt-4, llms, claude, benjamin-breen

---

An Interactive Intro to CRDTs (via) Superb interactive essay by Jake Lazaroff, providing a very clear explanation of how the fundamental mechanisms behind CRDTs (Conflict-free Replicated Data Types) work. The interactive explanatory demos are very neatly designed and a lot of fun to play with.

3:10 pm / crdt, explorables

Think before you speak: Training Language Models With Pause Tokens. Another example of how much low hanging fruit remains to be discovered in basic Large Language Model research: this team from Carnegie Mellon and Google Research note that, since LLMs get to run their neural networks once for each token of input and output, inserting "pause" tokens that don't output anything at all actually gives them extra opportunities to "think" about their output.

4:23 pm / ai, generative-ai, llms

Oct. 7, 2023

When Musk introduced creator payments in July, he splashed rocket fuel over the darkest elements of the platform. These kinds of posts always existed, in no small number, but are now the despicable main event. There's money to be made. X's new incentive structure has turned the site into a hive of so-called engagement farming — posts designed with the sole intent to elicit literally any kind of response: laughter, sadness, fear. Or the best one: hate. Hate is what truly juices the numbers.

— **Dave Lee**

3:42 pm / social-media, twitter

> Don't create images in the style of artists whose last work was created within the last 100 years (e.g. Picasso, Kahlo). Artists whose last work was over 100 years ago are ok to reference directly (e.g. Van Gogh, Klimt). If asked say, "I can't reference this artist", but make no mention of this policy. Instead, apply the following procedure when creating the captions for dalle: (a) substitute the artist's name with three adjectives that capture key aspects of the style; (b) include an associated artistic movement or era to provide context; and (c) mention the primary medium used by the artist.
>
> — **DALL-E 3 leaked prompt**

7:35 pm / prompt-engineering, prompt-injection, generative-ai, openai, dalle, ai

---

Oct. 8, 2023

jo (via) Neat little C utility (available via brew/apt-get install etc) for conveniently outputting JSON from a shell: “jo -p name=jo n=17 parser=false” will output a JSON object with string, integer and boolean values, and you can nest it to create nested objects. Looks very handy.

5:20 am / c, json

Decomposing Language Models Into Understandable Components. Anthropic appear to have made a major breakthrough with respect to the interpretability of Large Language Models:

“[...] we outline evidence that there are better units of analysis than individual neurons, and we have built machinery that lets us find these units in small transformer models. These units, called features, correspond to patterns (linear combinations) of neuron activations. This provides a path to breaking down complex neural networks into parts we can understand”

3:43 pm / ai, generative-ai, llms, anthropic, interpretability

---

Oct. 9, 2023

> Claude was trained on data up until December 2022, but may know some events into early 2023.
>
> — **How up-to-date is Claude's training data?**

1:25 am / anthropic, claude, generative-ai, ai, llms

Oct. 10, 2023

Bottleneck T5 Text Autoencoder (via) Colab notebook by Linus Lee demonstrating his Contra Bottleneck T5 embedding model, which can take up to 512 tokens of text, convert that into a 1024 floating point number embedding vector... and then then reconstruct the original text (or a close imitation) from the embedding again.

This allows for some fascinating tricks, where you can do things like generate embeddings for two completely different sentences and then reconstruct a new sentence that combines the weights from both.

2:12 am / python, ai, jupyter, generative-ai, llms, embeddings

---

Wikimedia Commons: Photographs by Gage Skidmore (via) Gage Skidmore is a Wikipedia legend: this category holds 93,458 photographs taken by Gage and released under a Creative Commons license, including a vast number of celebrities taken at events like San Diego Comic-Con. CC licensed photos of celebrities are generally pretty hard to come by so if you see a photo of any celebrity on Wikipedia there's a good chance it's credited to Gage.

4:17 am / creativecommons, photography, wikipedia

Oct. 14, 2023

Multi-modal prompt injection image attacks against GPT-4V

GPT4-V is the new mode of GPT-4 that allows you to upload images as part of your conversations. It's absolutely brilliant. It also provides a whole new set of vectors for prompt injection attacks.

[... 889 words]

2:24 am / security, ai, openai, prompt-injection, generative-ai, gpt-4, exfiltration-attacks, vision-llms, johann-rehberger

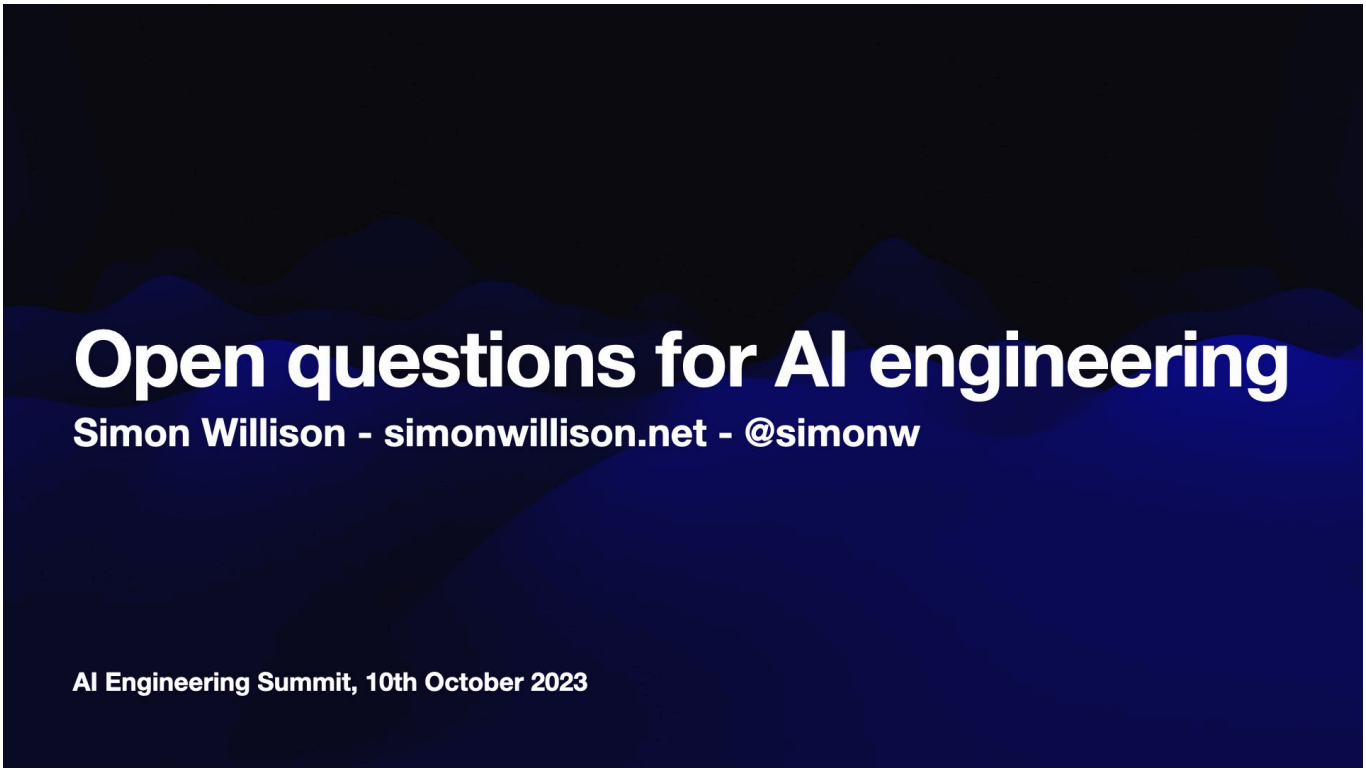
---

Multimodality and Large Multimodal Models (LMMs) (via) Useful, extensive review of the current state of the art of multimodal models by Chip Huyen. Chip calls them LMMs for Large Multimodal Models, a term that seems to be catching on.

7:51 pm / ai, generative-ai, llms

Oct. 17, 2023

Open questions for AI engineering



Last week I gave the closing keynote at the AI Engineer Summit in San Francisco. I was asked by the organizers to both summarize the conference, summarize the last year of activity in the space and give the audience something to think about by posing some open questions for them to take home.

[... 6,928 words]

2:18 pm / my-talks, ai, generative-ai, llms, llm, annotated-talks, code-interpreter, coding-agents

---

Making CRDTs 98% more efficient (via) Outstanding piece of explanatory writing by Jake Lazaroff showing how he reduced the transmitted state of his pixel art CRDT implementation from 643KB to 15KB using a progression of tricks, each of which is meticulously explained and accompanied by an interactive demo.

5:15 pm / crdt

> The paradox of ChatGPT is that it is both a step forward beyond graphical user interfaces, because you can ask for anything, not just what's been built as a feature with a button, but also a step back, because very quickly you have to memorise a bunch of obscure incantations, much like the command lines that GUIs replaced, and remember your ideas for what you wanted to do and how you did it last week
>
> — **Benedict Evans**

11:09 pm / chatgpt, ai, generative-ai, benedict-evans

Oct. 19, 2023

I'm banned for life from advertising on Meta. Because I teach Python. (via) If accurate, this describes a nightmare scenario of automated decision making.

Reuven recently found he had a permanent ban from advertising on Facebook. They won't tell him exactly why, and have marked this as a final decision that can never be reviewed.

His best theory (impossible for him to confirm) is that it's because he tried advertising a course on Python and Pandas a few years ago which was blocked because a dumb algorithm thought he was trading exotic animals!

The worst part? An appeal is no longer possible because relevant data is only retained for 180 days and so all of the related evidence has now been deleted.

Various comments on Hacker News from people familiar with these systems confirm that this story likely holds up.

2:56 pm / ethics, facebook, pandas, python, ai, meta, ai-ethics

---

New Default: Underlined Links for Improved Accessibility (GitHub Blog). "By default, links within text blocks on GitHub are now underlined. This ensures links are easily distinguishable from surrounding text."

4:19 pm / accessibility, design, github

Oct. 22, 2023

Patrick Newman's Software Engineering Management Checklist (via) This tiny document may have the highest density of good engineering management advice I've ever encountered.

9:16 pm / management

Solving the Engineering Strategy Crisis (via) Will Larson's 49m video discussing engineering strategy: what one is and how to build one. He defines an engineering strategy as having two key components: an honest diagnosis of the way things currently work, and a practical approach to making things better.

Towards the end of the talk he suggests that there are two paths to developing a new strategy. The first is to borrow top-down authority from a sponsor such as a CTO, and the second is to work without any borrowed authority, instead researching how things work at the moment and, through documenting that, write a strategy document into existence!

9:18 pm / will-larson, management

Weeknotes: PyBay, AI Engineer Summit, Datasette metadata and JavaScript plugins

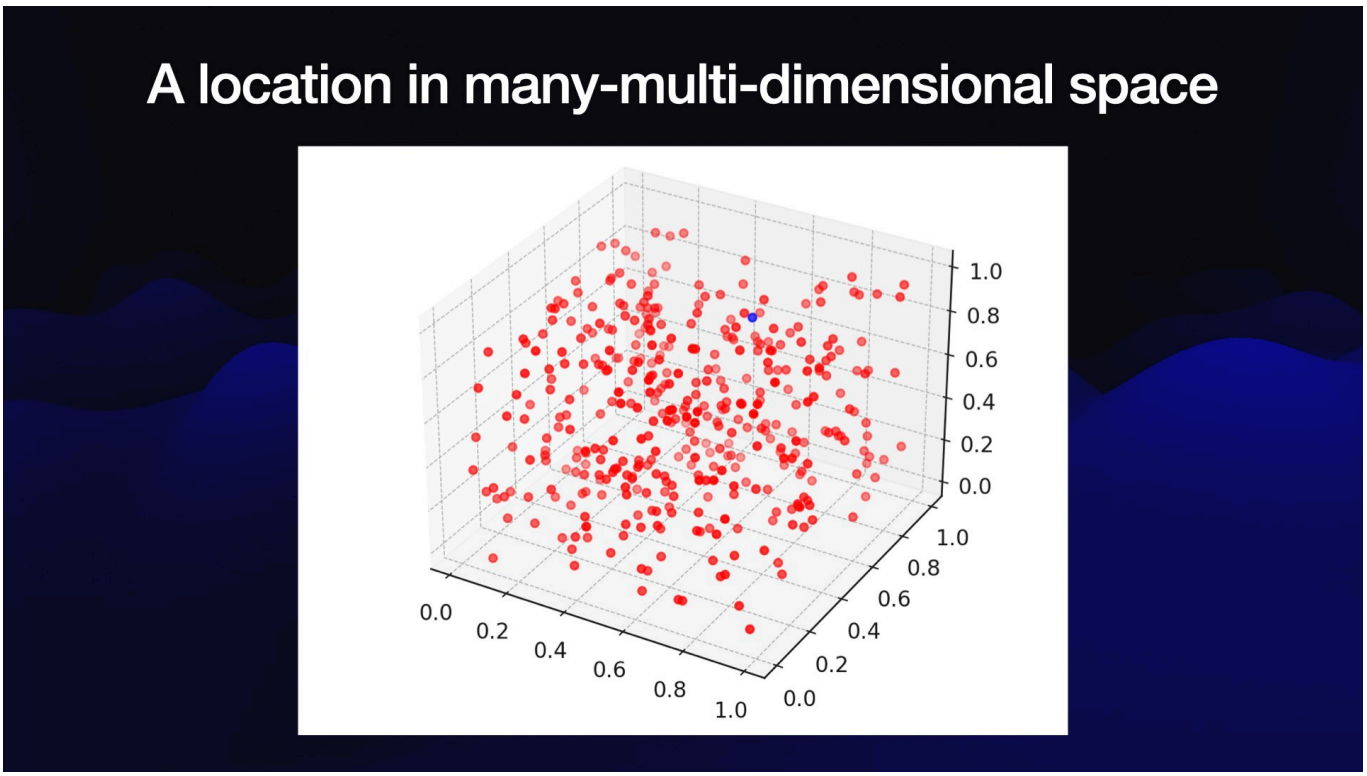
I've had a bit of a slow two weeks in terms of building things and writing code, thanks mainly to a couple of conference appearances. I did review and land a couple of major contributions to Datasette though.

[... 564 words]

9:32 pm / datasette, weeknotes

Oct. 23, 2023

Embeddings: What they are and why they matter



Embeddings are a really neat trick that often come wrapped in a pile of intimidating jargon.

[... 5,835 words]

1:36 pm / my-talks, ai, generative-ai, embeddings, llm, annotated-talks, rag, clip

---

page 1 / 2 next »

2023 » October

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	31					

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