

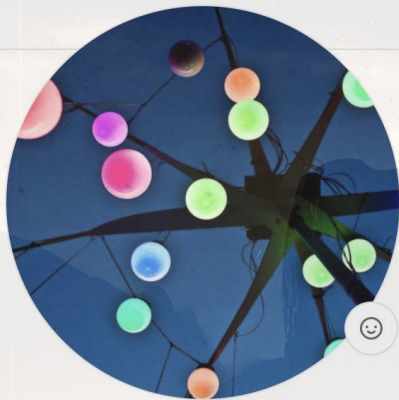


AI Reality Checkpoint

The Good, the Bad, and the Overhyped

Fall 2024





Maxime Beauchemin

mistercrunch

creator of Apache Airflow and Apache Superset - founder at Preset

Edit profile

🔔 1k followers · 11 following · ☆ 139

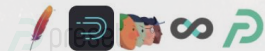
📁 preset-io

📍 San Mateo, CA

✉ maximebeauchemin@gmail.com

🔗 mistercrunch.blogspot.com

Organizations



- Passionate about building data tools!
- Started Apache **Airflow** at Airbnb in 2014
- Started Apache **Superset** at Airbnb in 2015
- Started **Preset** - The Apache Superset company in 2019



Apache
Airflow

preset

AI Reality Checkpoint: The Good, the Bad, and the Overhyped

- How did we get here!?
- The ASK-AI first reflex
- Use cases
 - Data-related workflows (data eng, data analysis, ...)
 - Programming
 - Founder / information worker-related
- Open source maintainer

- Gripes
- What's next!?

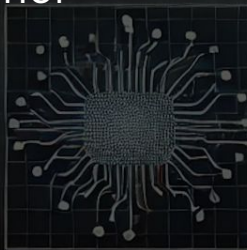
AI

AI



MOLIOAL

AGENDA

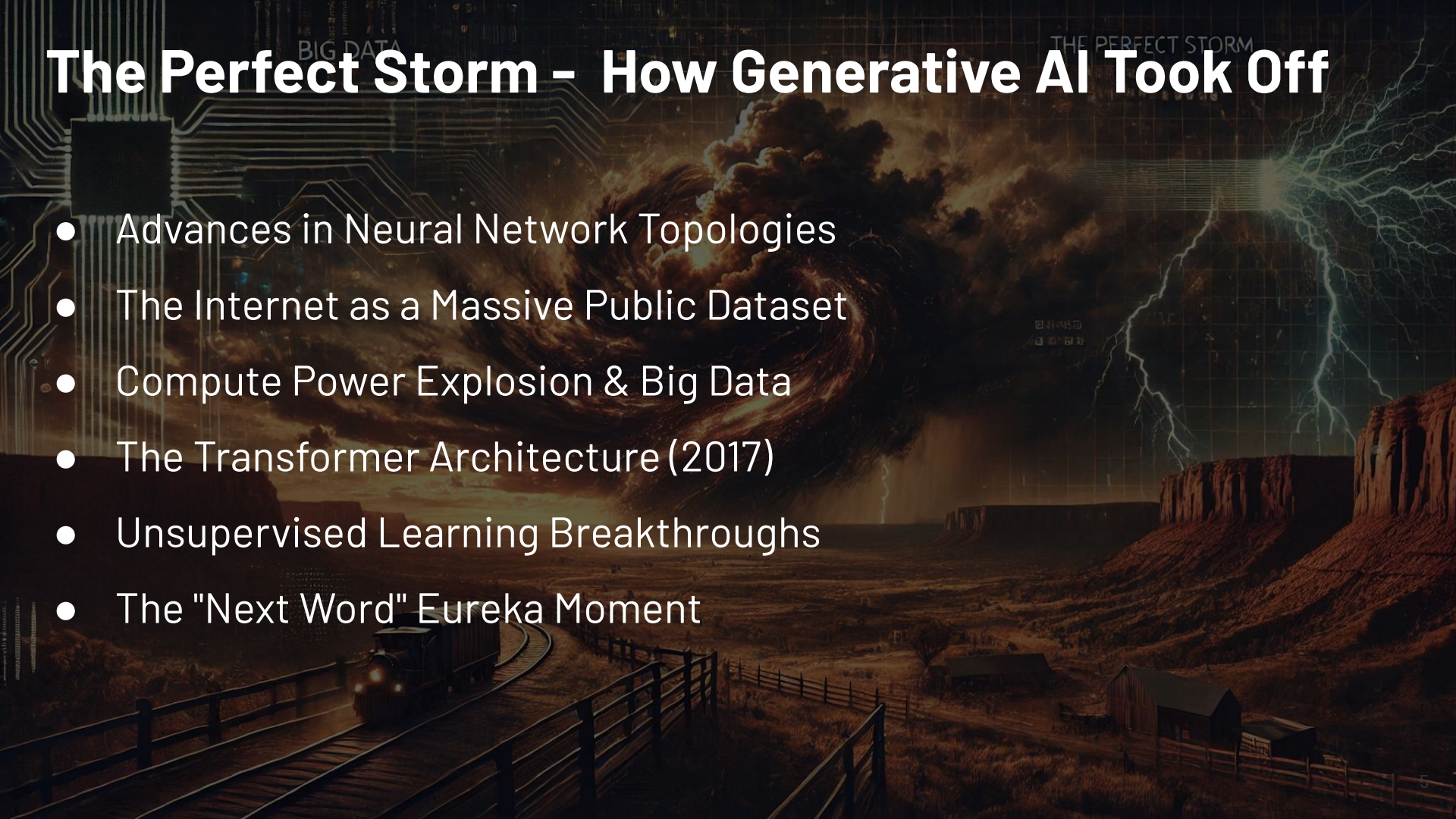


AI



The Perfect Storm – How Generative AI Took Off

- Advances in Neural Network Topologies
- The Internet as a Massive Public Dataset
- Compute Power Explosion & Big Data
- The Transformer Architecture (2017)
- Unsupervised Learning Breakthroughs
- The "Next Word" Eureka Moment





THE GOOD OF THE UGLY

... open source as a leader right out of the gate ...

A still life photograph with a vintage, industrial aesthetic. In the center-left, a glowing spherical light source, possibly a lamp or a light bulb, casts a warm, golden light. To its left is a vintage portable radio with a speaker grille and several knobs. In the foreground, a small, detailed steam engine with a large flywheel and a chimney is positioned. To the right of the steam engine, a handgun is visible. The background is dark and textured, with a framed picture or poster hanging on the wall. The overall mood is nostalgic and mysterious.

USE CASES

THE GOOD, AND THE UGLY

USE CASES

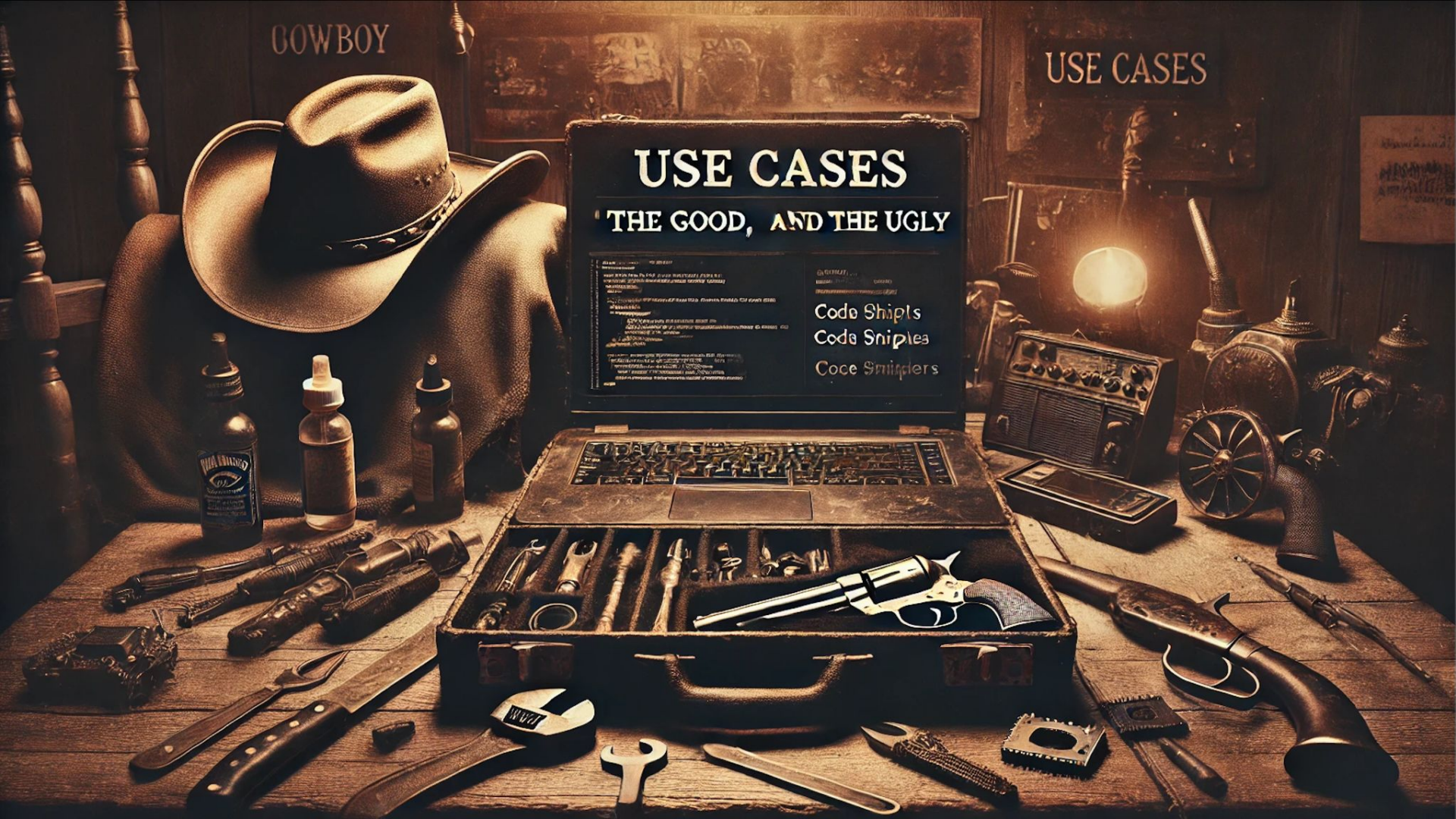
THE GOOD, AND THE UGLY

Code Snipits

Code Sniples

Coce Sniptlers

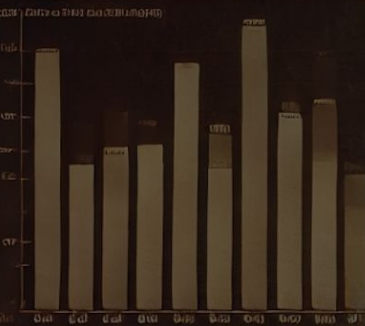
Code Shipls
Code Sniples
Coco Smimplers



Good bad

A man in a cowboy hat and poncho, holding a lasso, with the text "As a Data Practitioner" overlaid. The background is a textured, brownish-gold surface with faint, vertical text on the right side that reads "The Good, the Bad, and the Ugly".

DAVA



As a Data Practitioner



Good

- AI assist - SQL authoring
- Scripting acceleration
- As an advisor on approach
- Suggesting code snippet
- Ideating / advising on KPIs, industry standards
- Abstract modeling questions



Bad

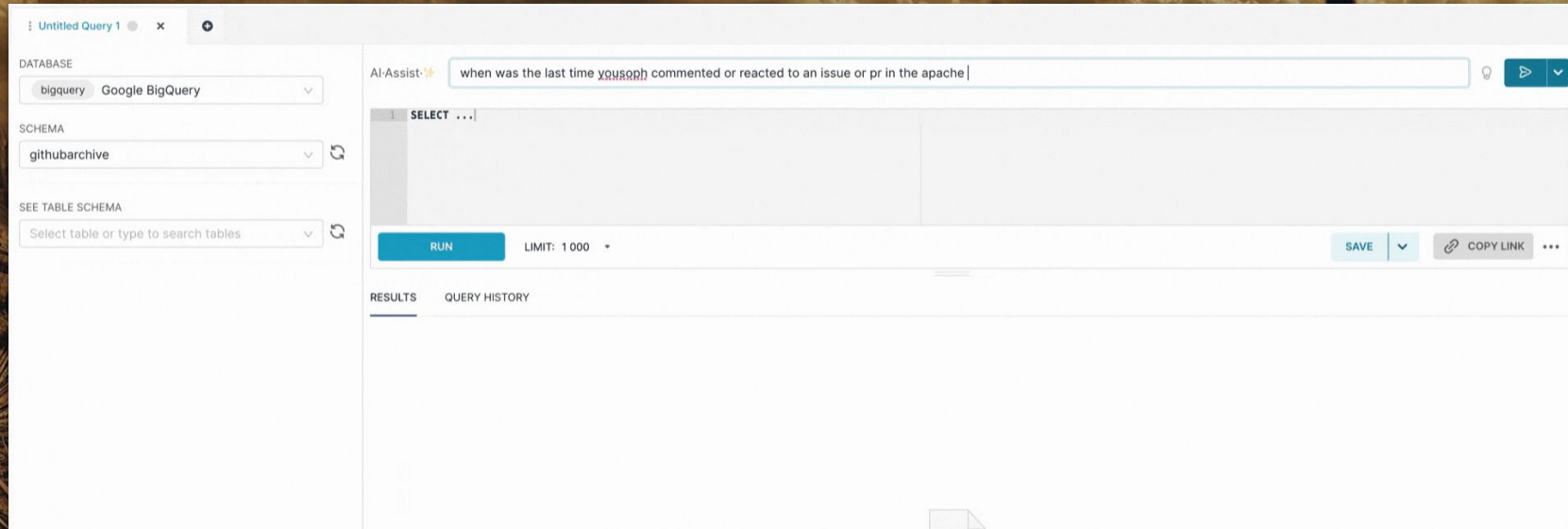
- Anything that requires deep or wide context



Overhyped

- the "Analyst Agent" idea
- RAG - making progress...
- Code Interpreter (hit or miss)
- Anything that requires executive skills

Example: authoring SQL with AI assist in Preset.io



The screenshot displays the Preset.io web interface for authoring SQL queries. The interface is divided into a left sidebar and a main workspace.

Left Sidebar:

- Untitled Query 1**: Tab title with a close button.
- DATABASE**: A dropdown menu showing **bigquery** and **Google BigQuery**.
- SCHEMA**: A dropdown menu showing **githubarchive**.
- SEE TABLE SCHEMA**: A dropdown menu with the text "Select table or type to search tables".

Main Workspace:

- AI-Assist**: A text input field containing the prompt: "when was the last time yousoph commented or reacted to an issue or pr in the apache". To the right of the input are a lightbulb icon and a dropdown arrow.
- SQL Editor**: A large text area showing the generated SQL query: `1 SELECT ...`.
- Buttons**: A blue **RUN** button, a **LIMIT: 1 000** dropdown, a **SAVE** button, and a **COPY LINK** button.
- Navigation**: Tabs for **RESULTS** and **QUERY HISTORY**.

Example: Cloning table structures into another schema

in bigquery, I'd like to create/clone all the table structures from a dataset to another, what's the easiest way to do this?



To clone the table structures (without the data) from one dataset to another in BigQuery, you can follow these steps:

1. List all tables from the source dataset:

You can get the list of tables in the source dataset using the `bq` command-line tool:

```
bash
```

[Copy code](#)

```
bq ls --format=prettyjson [project_id]:[source_dataset_id]
```

This will list all the tables in the source dataset.

2. Use `bq mk` to create the new tables in the target dataset:

For each table, you can copy the schema and create an empty table in the new dataset.

Here's a simple example using `bq`:

```
bash
```

[Copy code](#)

```
bq mk --table [project_id]:[target_dataset_id].[table_name] [project_id]:[source_dataset_id]
```


Example: Assisted me in writing “entity-centric data modeling” blog



DATA ENGINEERING

Introducing Entity-Centric Data Modeling for Analytics

Maxime Beauchemin April 06, 2023



Subscribe

Context

In this blog post, I am introducing **entity-centric data modeling (ECM)**, a new approach to data modeling for analytics that elevates the core idea of an “entity” (ie: user, customer, product, business unit, ad campaign, etc) at the very top of things. Many of the ideas here are not new and are inspired by combining concepts from [dimensional modeling](#) and [feature engineering](#) - pushing to **anchor on “entities”** and to **actively bring metrics** and **more complex data structures inside dimensions**.

THE GOOD, THE BAD THE UGLY

SOFTWARE

As a Software Engineer



As a Software Engineer



Good

- Autocomplete! GH copilot & Codeium plugins
- Functional programming (over object orient)
- Refactoring! js -> ts, type hints, migrations
- Unit tests generation
- Jinja templates for prompts
- Naming things



Bad





- Anything that requires deep or wide context
- API usage - lots of hallucinations - unpredictable
- OOP - inheritance





Overhyped

- the "Programmer Agent" idea
- Anything that requires executive skills

Wrote 80%+ of the README and code for Promptimize

 README  Apache-2.0 license  


 **promptimize!** 

License

Apache 2.0

pypi package

0.2.3

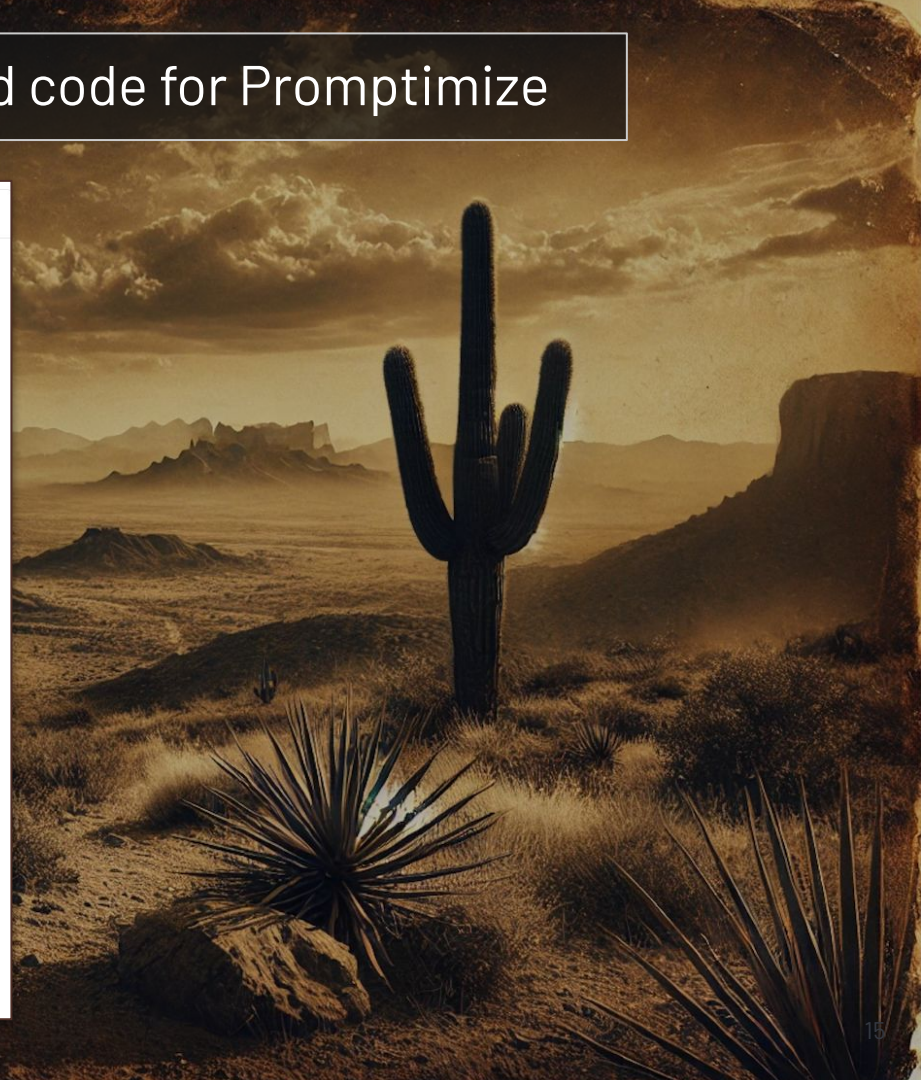


Promptimize is a prompt engineering **evaluation** and **testing** toolkit.

It accelerates and provides structure around prompt engineering at scale with confidence, bringing some of the ideas behind test-driven development (TDD) to engineering prompts.

With promptimize, you can:

- Define your "prompt cases" (think "test cases" but specific to evaluating prompts) as code and associate them with evaluation functions
- Generate prompt variations dynamically
- Execute and rank prompts test suites across different engines/models/temperature/settings and compare results, bringing the hyperparameter tuning mindset to prompt engineering
- Get reports on your prompts' performance as you iterate. Answer question around how different prompt suites are performing against one-another. Which individual cases or categories of cases improved? regressed?



A man in a cowboy hat and vest holding a tablet, with the text 's a Startup Founder' overlaid. The background is a textured, brownish-gold surface with faint, illegible text and a grid pattern on the right side.



As a Startup Founder

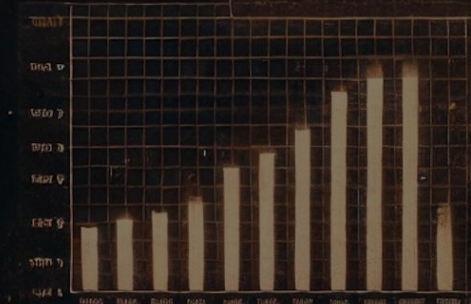
PRODUCT RADOOPS

STARTUP
WESTERN
WESTERN

← GOOD-BAD

PITCH DECK

PRODUCT



As a Startup Founder



Good

- Product marketing
- Judgment-free thought partner / advisor
- Content generation
- Comms / advice
- Legal / advice



Bad

- Running your company



Overhyped

- Anything that requires executive skills
- Staffing - AI won't replace 80% of your opex, at least not just yet

Example: Crafting a customer email

hey I have some evidence that some of our paid customers are user service accounts and therefore breaking our TOS, can you help me write an email to the account owner?



Sure thing. Here's a draft for your email:

Subject: Clarification Needed on Account Usage

Hi [Customer Name],

I hope this email finds you well. After reviewing your account activity, we've noticed some indications that it might be used by service accounts, which falls outside the scope of our terms of service for paid users.

To ensure we're aligned, can you confirm whether any of your seats are being used by non-human or automated accounts? If so, we'll need to discuss how to bring the account into compliance with our policies.

Please let me know if you'd like to set up a quick call to go over any questions or clarify further.

Thanks for your attention to this,
Max



The Good, and the Ugly;

Open Source

Open Source Maintainer

Open Source

GitHub Pull Requests

GitHub Pull Requests

GitHub
Code Reviews

Comi

As an Open Source Maintainer / GitHub User

Open Source Maintainer

GitHub 'Open Source Maintainer'

GitHub > Open Source
606.6500 > Open Source
606.6500 > Open Source
606.6500 > Open Source
606.6500 > Open Source

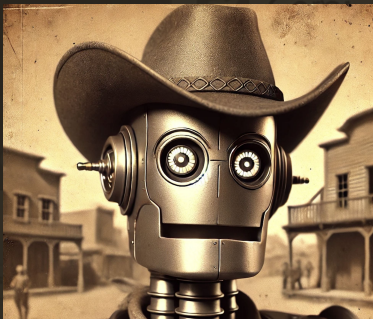
Gode Shup

Open Source Maintainer

As an Open Source Maintainer

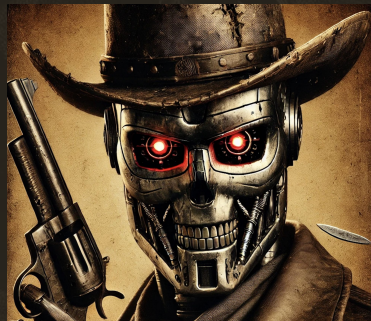
Good

- Bots



Bad

- Bots



Overhyped

- Bots



Example: Sensitive / thoughtful community comms



hey! I just got CCed as the chair of Apache Superset for what seems to be a geo-political issue in relation to the maps we exposed in Superset. Note that I have no opinion on the matter, and would rather stay out of any political topic as an open source maintainer. I'm worried that our actions, whatever they may be, or inaction may be interpreted as taking as side. I'd love to push on the Apache Software Foundation to provide clear guidelines and handle communications around such topics. Here's the related thread, please provide email draft for me answer to the internal communication:

Original message to the apache mailing list:

Subject: Re: Map-related geopolitical concerns in Superset

Hi DW,

Thanks for reaching out and for your thoughtful approach to this issue.

As an open-source project, Apache Superset aims to be a neutral platform that serves a global audience. Given the sensitivity and complexity surrounding geopolitical topics like this, I believe it's crucial for us to avoid taking any particular stance that could be interpreted as political.

This is a space where clear guidance from the Apache Software Foundation (ASF) would be incredibly valuable. I would suggest that ASF consider establishing a set of guidelines or protocols to help projects like Superset navigate these geopolitical challenges. It would help ensure that we remain aligned with the Foundation's values and avoid unintentionally stepping into political disputes.

In this specific case, I think it's important to clarify the source of the maps we are using and ensure they follow some globally recognized standards, such as those provided by the UN. That way, we can stay focused on providing a technical solution while maintaining neutrality.

I'd love to see ASF take the lead on managing communications around these sensitive issues, ensuring that projects like ours can continue to focus on their technical mission without being drawn into political debates.

Happy to discuss further if needed, and let me know if you need anything else from me to help guide the next steps.

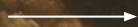
Best regards,

Max



Dosu AI (dosu.dev)

Intricate issue



apache / superset

Code Issues 622 Pull requests 317 Discussions Actions Projects 6 Wiki Security 36 Insights

flask_log.current_user set to None in SQL_QUERY_MUTATOR while using SQLLab #30168

Closed 3 tasks done dmarkey opened this issue 4 days ago · 3 comments

dmarkey commented 4 days ago

Assignees
No one—assign yourself

Bug description

While implementing the correct...

dosubot bot commented 4 days ago

Hey there, @dmarkey! I'm here to help you with any bugs, questions, or contributions you have for Apache Superset. Let's tackle this issue together!

The issue you're encountering with `flask_login.current_user` being `None` during the execution of `SQL_QUERY_MUTATOR` in SQLLab sessions is due to differences in the execution context between SQLLab and navigating graphs. In SQLLab sessions, the user ID is explicitly retrieved and stored in the `SqlJsonExecutionContext` class using the `get_user_id()` function, rather than relying on `flask_login.current_user` directly [1].

To work around this, you can modify your `SQL_QUERY_MUTATOR` to use the `get_user_id()` function to retrieve the user ID. Here's an example:

```
from superset.utils.core import get_user_id

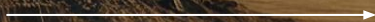
def SQL_QUERY_MUTATOR(sql: str, **kwargs) -> str:
    user_id = get_user_id()
    print(user_id) # This should print the correct user ID during a SQLLab session
    return sql
```

This approach ensures that you get the correct user information even in the SQLLab context.

To continue talking to Dosu, mention @dosu.

3 1

Intricate answer



Success!



dmarkey commented 4 days ago

That actually works.. first time I've been impressed by AI.

2 1

Kapa.ai + RunLLM

(as featured on our Slack's #ask-ai)

Here:

- 2 bots offer different answers
- Long conversations ensue

achieve that?



37 replies Last reply 3 days ago



ZHE CHONG LOW 9:03 PM

how to enable CSV upload in superset



9 replies Last reply 4 days ago

Thread



Manish Thursday at 7:03 AM

what is Other option in supported database?



4 replies



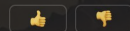
ThreadoAI APP Thursday at 7:03 AM

Hey @Manish

Superset supports a wide range of databases, including any SQL-speaking datastore or data engine that has a Python DB-API driver and a SQLAlchemy dialect. Some of the major supported databases include:

- StarRocks
- Snowflake
- SQLite
- SQL Server
- Teradata

[See more](#)



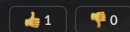
Tag me @threadoai to ask any follow up questions, always happy to assist.



RunLLM APP Thursday at 7:03 AM

The "Other" option in the supported databases allows users to connect to databases that are not explicitly listed in the predefined options. This option typically uses a SQLAlchemy URI for the connection. It provides flexibility for connecting to various databases by manually specifying the connecti...

[See more](#)



Hint: Mention @RunLLM in the thread for followups.



kapa.ai APP Thursday at 7:03 AM

Hi @Manish. I'm kapa.ai, your friendly support bot. I will try to answer your question by searching through the knowledge base. 🙌

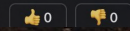
Feel free to ask a follow-up question once I've answered by mentioning me with @kapa.ai

Apache Superset supports a wide range of databases. If your database or data engine isn't on the list but a SQL interface exists, you can file an issue on the [Superset GitHub repo](#), so the team can work on documenting and supporting it.

The main criteria for a database to be supported by Superset is...

[See more](#)

kapa.ai is still learning and improving, please let me know how I did by reacting below



Kapa.ai – i18n!

(as featured on our Slack's #ask-ai)

0

Oguzhan Deniz Saturday at 4:15 PM

Jinja template ile iki farklı veri setinde bulunan aynı veriyi tek bir dashboardda nasıl filtreleme işlemi yapabilirim ?

4 replies

t.

ThreadoAI APP Saturday at 4:15 PM

Hey @Oguzhan Deniz

To filter the same data from two different datasets in a single dashboard using Jinja templates, you can follow these steps:

1. Create Filters in the Dashboard:

- Start by adding filters to your dashboard. You can do this by selecting the right arrow to expand the Filter area and then choosing "+ Add/Edit Filters."
- In the "Add and edit filters" window, enter a name for your filter and select the appropriate dataset and column.

2. Use Jinja Templating:

- Jinja templating allows you to create dynamic queries that can handle complex filtering logic. You can use Jinja to ensure that the filters apply correctly across different datasets.
- For example, you can create a filter that applies an OR condition instead of an AND condition by using Jinja syntax in your query.

3. Map Filters to Charts:

- Ensure that the filters are correctly mapped to the charts. You can do this by selecting the "Scoping" tab in the filter settings and mapping the filter to the relevant charts.
- If you have charts from different data sources, make sure each chart's dataset is configured to accept the filter.

4. Test the Filters:

- After setting up the filters, test them to ensure they work as expected. Apply the filters and check if the charts update correctly with the filtered data.

For more detailed guidance on using Jinja templating in Apache Superset, you can refer to these resources:

- [Superset SQL Templating](<https://superset.apache.org/docs/installation/sql-templating/>)
- [Intro to Jinja Templating in Apache Superset](<https://preset.io/blog/intro-jinja-templating-apache-superset/>)

I hope this helps! If you have any further questions, feel free to ask. 😊

Sources :

1. Intro To Jinja Templating in Apache Superset™ | Preset

See less

Dosu



dosubot bot commented on Jul 12

Hey [@JatinVenkatesh](#)! I'm here to assist you with any bugs, questions, or contributions you have while we wait for a human maintainer to look into your issue.

The `DISALLOWED_SQL_FUNCTIONS` error in Apache Superset version 4.0.2 is due to the introduction of a setting that restricts the use of certain SQL functions. This setting is used to enhance security by disallowing unsafe SQL functions.

To resolve this issue, you can modify the `DISALLOWED_SQL_FUNCTIONS` setting in your Superset configuration. Here is an example of how to adjust this setting:

```
DISALLOWED_SQL_FUNCTIONS: dict[str, set[str]] = {
    "postgresql": {"version", "query_to_xml", "inet_server_addr", "inet_client_addr"},
    "clickhouse": {"url"},
    "mysql": {"version"},
}
```

You can remove or modify the functions listed under `"postgresql"` to fit your requirements. For instance, if you need to allow the `version` function, you can update the configuration as follows:

```
DISALLOWED_SQL_FUNCTIONS: dict[str, set[str]] = {
    "postgresql": {"query_to_xml", "inet_server_addr", "inet_client_addr"},
    "clickhouse": {"url"},
    "mysql": {"version"},
}
```

Additionally, ensure that your database permissions and configurations are correctly set to allow the execution of the required SQL functions.

For more details, you can refer to the relevant configuration documentation [\[1\]](#).

To continue talking to [Dosu](#), mention [@dosu](#).



michael-s-molina commented on Jul 17 • edited • Member

Wow. [@dosubot](#) nailed! [@rusackas](#)



PhilBlackRose commented on Aug 6

Author

Thanks bot, updating did indeed clear the error.



jvines commented 2 weeks ago

Contributor

Hi, sorry it took me so long to get to this. I'll open a PR to solve this following Dosu's suggestion.



PhilBlackRose closed



dosubot bot commented on Aug 6

Great to hear that updating resolved the error!
Happy coding!

To continue talking to [Dosu](#), mention [@dosu](#).



PhilBlackRose commented on Aug 6

Author

Thanks bot, updating did indeed clear the error.



PhilBlackRose closed this as [completed](#) on Aug 6



dosubot bot commented on Aug 6

Great to hear that updating resolved the error! If you encounter any more issues or have further questions, feel free to ask.
Happy coding!

To continue talking to [Dosu](#), mention [@dosu](#).



This approach ensures that the Gunicorn process cannot exceed the specified memory limit, helping to prevent crashes due to memory exhaustion.

To continue talking to [Dosu](#), mention [@dosu](#).



Habeeb556 commented 5 days ago

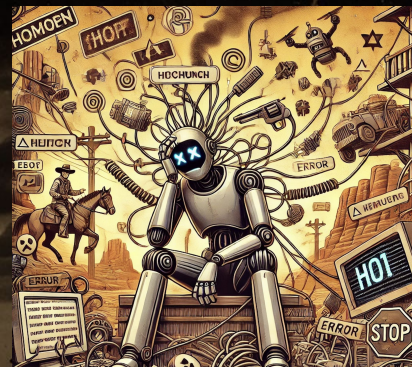
Author

Awesome!



Gripes!

- Is GPT getting stupider over the 4-o cycle?
- GPT is incredibly verbose - STFU!
- AI always one-shots the answer, never asks for clarification
- Hallucination - guesses how APIs should work
- Don't answer from weights, search the damn web!



Wrapping up with some related thoughts

- This is still the first inning - this talk is already obsolete!
- Intelligence / computation / energy limits? No problem.
- RAG/integration implementation are promising but clunky
- Trust still an issue - work needs expert validation
- Great for creative workflows and ideation - be careful where correctness matters
- Highly encourage people to develop their "first reflex"

A vintage movie theater marquee with a circular frame and curtains. The marquee is made of wood and has a circular opening in the center. The opening is framed by several concentric rings. The curtains are made of a heavy, brown fabric and are tied back with thick, braided ropes. The marquee is decorated with various items, including a clock on the left, a small sign on the right, and a large, stylized sign in the center. The overall appearance is that of a classic, old-fashioned movie theater.

*That's All
Folks!*

WOSLEY
TUNES