

DevOps Experiments with AI



Time to Experiment

Prepared by: Daniel Flower

Date: 25/11/2023

Introduction

This is a talk about using and coding with Large Language Models.
I believe that these are very powerful, but no one knows their full potential or
how to maximise their use yet.

Therefore, now is a time to try many experiments to see what works.

This talk has a very brief overview of Large Language Models and then just
shows a few experiments.

The intention is give ideas about the kinds of things we can do with it, to inspire
others to start experimenting more.

Large language models

How they work

- It takes some words as input. It generates the next word. It repeats this to make sentences and paragraphs.
- It remembers nothing. It learns nothing. So we need to give it everything it needs to know as input.
- It tries to guess what "sounds" best - not what is true.
 - it is able to guess based not just on stats, but by generalising, "reasoning", and other things

```
function guessNextWord( input: words Array<String> ) output: String {  
    var numbers = convertWordsToNumbers(words);  
    var nextNumber = guessNextNumber(numbers);  
    return convertNumberToWord(nextNumber);  
}
```

```
function guessNextNumber( input: numbers Array<Number> ) output: Number {  
    if (numbers.size > 8192) throw "Sorry! Too many inputs";  
    var guess = numbers[0] * 0.123124 + numbers[1] * 0.584935 + ...  
    // Huge amount of adding and multiplying numbers, but no state nor network calls etc  
    // Training the model takes a huge amount of time and money  
    return guess;  
}
```

Environment monitoring

Monitoring disk usage

Idea: every 10 minutes feed the model some monitoring info. Ask it to highlight any issues.

The prompt: You are a linux administrator, monitoring disk usage. You should look out for situations where a disk mount is increasing and if you think there might be an issue give a estimate of how long until the disk space is used. The user will supply disk usage stats along with the date - you should comment each time on the situation.

```
Filesystem      Size  Used Avail Use% Mounted on
devtmpfs        566G     0  566G   0% /dev
tmpfs           567G   792K  567G   1% /dev/shm
tmpfs           567G   4.0G  563G   1% /run
/dev/disk1      24G    9.8G   13G  43% /
/dev/disk2     192G    9.4G  173G   6% /opt
```

```
As per the output of the 'df -h' command at 16:30, the usage of '/dev/disk1' has increased further to 51% from the earlier 47% noted at 16:20. Given that this is a noticeable increase in usage during the elapsed 10 minutes since the previous reading and may indicate that the disk space usage is increasing rapidly. Based on this pattern, it's anticipated that the disk space usage will be full in approximately 20 minutes from the current time. It's suggested to investigate the root cause of the increased disk usage for '/dev/disk1' and take appropriate actions accordingly to avoid unforeseen downtimes.
```

Monitoring GC logs

Idea: detect memory issues by looking at GC logs

Input: Tell me about the behaviour of a java app based on the gc logs below which has Xmx256m in terms of how memory usage looks and if there are any issues or recommendations. Respond in both English and Chinese.

```
[2023-11-18T15:34:19.137+0800][info][gc] Using Serial
[2023-11-18T15:34:19.505+0800][info][gc] GC(0) Pause Young (Allocation Failure) 7M->2M(27M) 10.936ms
[2023-11-18T15:34:19.696+0800][info][gc] GC(1) Pause Young (Allocation Failure) 9M->3M(27M) 10.975ms
[2023-11-18T15:34:19.856+0800][info][gc] GC(2) Pause Young (Allocation Failure) 11M->5M(27M) 8.755ms
[2023-11-18T15:34:20.068+0800][info][gc] GC(3) Pause Young (Allocation Failure) 12M->6M(27M) 5.711ms
[2023-11-18T15:34:20.287+0800][info][gc] GC(4) Pause Young (Allocation Failure) 13M->6M(27M) 6.002ms
[2023-11-18T15:34:20.453+0800][info][gc] GC(5) Pause Young (Allocation Failure) 14M->9M(27M) 14.697ms
[2023-11-18T15:34:20.683+0800][info][gc] GC(6) Pause Full (Metadata GC Threshold) 17M->10M(27M) 31.128ms
[2023-11-18T15:34:31.784+0800][info][gc] GC(7) Pause Young (Allocation Failure) 17M->11M(27M) 4.372ms
[2023-11-18T15:34:32.204+0800][info][gc] GC(8) Pause Young (Allocation Failure) 18M->11M(27M) 5.303ms
[2023-11-18T15:34:32.466+0800][info][gc] GC(9) Pause Young (Allocation Failure) 19M->12M(27M) 4.589ms
[2023-11-18T15:34:32.608+0800][info][gc] GC(10) Pause Young (Allocation Failure) 19M->12M(27M) 3.784ms
[2023-11-18T15:34:32.956+0800][info][gc] GC(11) Pause Young (Allocation Failure) 19M->12M(27M) 7.896ms
[2023-11-18T15:34:33.592+0800][info][gc] GC(12) Pause Young (Allocation Failure) 20M->13M(27M) 7.678ms
```

Monitoring GC logs

Idea: detect memory issues by looking at GC logs

Based on the GC logs provided, the Java app's memory usage is experiencing frequent young generation garbage collection due to allocation failures. The application is using the serial garbage collector. Each pause time is relatively low, indicating that the application is not experiencing significant issues related to garbage collection. However, the fact that allocation failures trigger frequent garbage collection indicates that there may be memory leaks or inefficiencies in the application's memory usage. The app is configured with `Xmx256m`, which limits the heap size to 256 MB. This may be a limiting factor in the application's performance if it requires more memory than this limit allows. It is recommended to analyze the application's memory usage and optimize it to minimize unnecessary object creation and maximize memory reuse.

根据提供的 GC 日志，Java 应用程序的内存使用率正经历频繁的 young generation 垃圾回收，这是由于分配失败导致的。应用程序正在使用串行垃圾收集器。每次暂停时间相对较低，表明应用程序没有遇到与垃圾回收相关的重大问题。然而，分配失败引发频繁垃圾回收的事实表明，应用程序的内存使用可能存在内存泄漏或不足的情况。该应用程序配置了 `Xmx256m`，将堆大小限制为 256 MB。如果应用程序需要比此限制更多的内存，这可能是应用程序性能的限制因素。建议分析应用程序的内存使用情况并进行优化，以最小化不必要的对象创建并最大程度地重用内存。

Any potential?

Some things work well; some things don't.

This is overkill to just monitor disk usage. But give it more data and it should be able to spot things that would be difficult to monitor for up-front, for example it might notice relationships between a process and its thread, CPU, memory or disk usage.

A lot more experimentation is needed here.


Security reviews

Looking at Git diffs

Finding security issues in code reviews

Can a language model find issues in code changes?

```
const qs = new URLSearchParams(location.search);
const address = qs.get('address');
const url = qs.get('url');
document.querySelector('h1').textContent = address || 'Error: no address s
fetch('/udn/api/v1/instances/' + encodeURIComponent(address) + '/files?fie
  .then(r => {
    if (!r.ok) throw 'Got a ${r.status} fr instance API lookup';
    return r.json();
  })
```



```
document.querySelector('h1').textContent = address
fetch('/udn/api/v1/instances/' + address + '/files
  .then(r => {
```

How to do it

A code base is too large to put into a language model prompt, but what about a git diff?

```
You are an expert software security reviewer. Please succinctly summarise the
following git diff and if focus on any security issues or bugs it may introduce:

...

diff --git a/src/main/resources/web/file-list.module.js b/src/main/resources
/web/file-list.module.js
index 9dc5006..42e90a6 100644
--- a/src/main/resources/web/file-list.module.js
+++ b/src/main/resources/web/file-list.module.js
@@ -14,7 +14,7 @@ const FilesList = () => {
    const address = qs.get('address');
    const url = qs.get('url');
    document.querySelector('h1').textContent = address || 'Error: no address
specified';
-    fetch('/udn/api/v1/instances/' + encodeURIComponent(address) +
'/files?field=NAME&field=SIZE&field=RELATIVE_PATH&field=LAST_MODIFIED&field=URL')
+    fetch('/udn/api/v1/instances/' + address + '/files?field=NAME&field=SIZE&
field=RELATIVE_PATH&field=LAST_MODIFIED&field=URL')
      .then(r => {
        if (!r.ok) throw `Got a ${r.status} from instance API lookup`;
        return r.json();
      })
...


```

Example response

The change is summarised and potential issue highlighted

The git diff displays a single change in a JavaScript file (`file-list.module.js`). In this change, the `encodeURIComponent()` function is removed from the `address` query parameter when making a `fetch()` call to the API endpoint.

The removal of `encodeURIComponent()` poses a potential security risk as it allows unsafe characters to be passed to the API. This can lead to bugs and security vulnerabilities such as Cross-site Scripting (XSS) or API service issues.

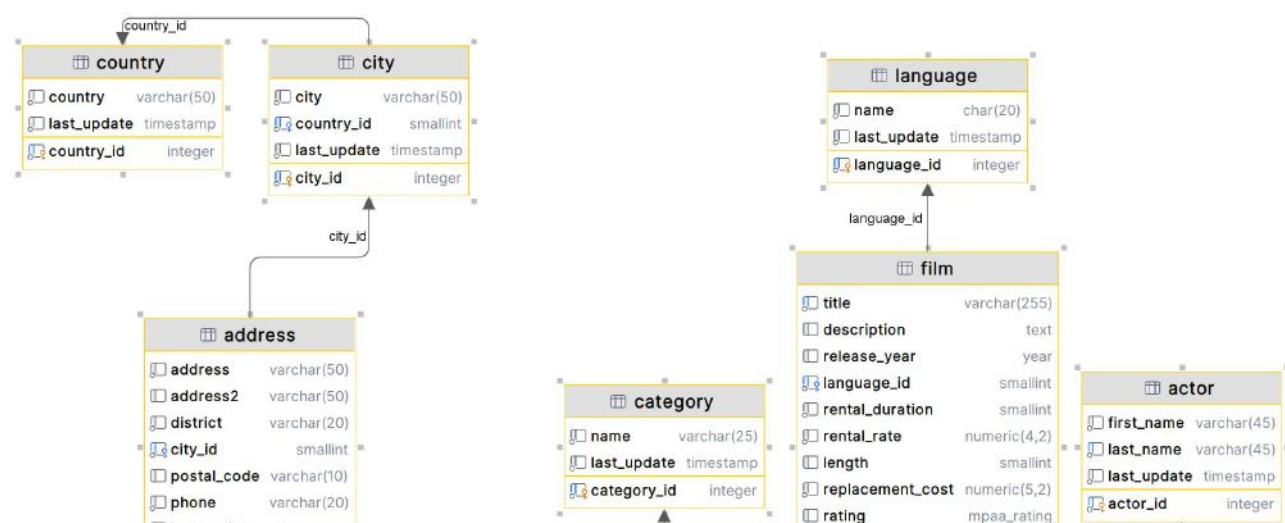
Possibilities

- Integrate security code review into build pipeline
- May have many false positives and false negatives
- Could go beyond security: e.g. generate release notes from code diffs

Talking to a database

SQL and JavaScript code generation

Example database




```
SELECT * FROM film;
```

dvdrental.public.film x Tx ✓ ↺

of 501+ > < ↺ ⌂ + - ↶ ↷ Tx: Auto DDL ↗

film_id	title	release_year	description
37	Arizona Bang	2006	A Brilliant Panor
354	Ghost Groundhog	2006	A Brilliant Panor
144	Chinatown Gladiator	2006	A Brilliant Panor
309	Feud Frogmen	2006	A Brilliant Refle

 Film DB是一个包含多个表格的PostgreSQL 14数据库，这些表格与电影及其租借有关。这些表格包括演员、类别、国家、城市、地址、客户、语言、电影、电影演员、电影类别、库存、员工、租赁、支付和商店。这些表格包含有关演员、电影类别、国家、城市、租赁客户、语言、电影、电影类别、库存、商店员工、租赁交易以及租赁付款信息的信息。表格通过各种外部键和主键连接在一起以显示数据之间的关系。该数据库允许存储和检索与电影及其租借历史有关的信息。

Trying it out...



What can you do? (answer in Chinese)



我可以查询电影数据库中的表格，并以自由文本或者执行SQL查询的方式回答用户的问题。用户可以提问任何关于电影库的问题。



Show me some film titles with release year and category



以下是电影标题、发行年份和类别信息的查询结果：

release_year	name	title
2006	Action	Academy Dinosaur
2006	Horror	Ace Goldfinger
2006	Documentary	Adaptation Holes

Performance Testing

Will this understand us and give us good information?



How much to rent this movie?



To find the rental rate for Academy Dinosaur, you can run the following query:

rental_rate
0.99

This will give you the rental rate for the movie, which is a numeric value with two decimal places.



Prompt engineering

Giving the model the right information

The maximum size of input to a large language model is typically around 5 to 20 thousand words. A database might have hundreds of millions of records. So how is it possible to use a large language model to answer questions about a database?

```
conversation.messages.push({role: "system", content: `You are a query engine for a database called Film DB. The user will ask you questions. You may answer in free text or by running a query in the DB. Any sql you write will be executed and the user will see the results of that query in a DB table. Please always add LIMIT 100. Below is the DB schema for the PostgreSQL 14 database:
```

```
Remember: the user will see the result of the SQL call, not the SQL statement itself. So any commentary should assume the user is reading a table of data rather than a SQL statement. For example do not say "please run the following query to see the results:" - instead say "here are your results: \\\`sql ...etc"
```

```
create table actor|
(
  actor_id integer not null primary key,
  first_name varchar(45) not null,
  last_name varchar(45) not null,
  last_update timestamp not null
);
```

```
create table category
(
  category_id integer not null primary key
```

Preparing the prompt

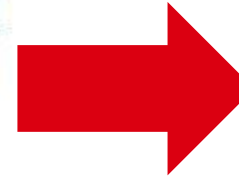
Simplify for the LLM

```
create table if not exists public.actor
(
    actor_id integer default nextval('actor_actor_id_seq'::regclass) not null
        primary key,
    first_name varchar(45)
        not null,
    last_name varchar(45)
        not null,
    last_update timestamp default now()
        not null
);

alter table public.actor
    owner to postgres;

create index if not exists idx_actor_last_name
    on public.actor (last_name);
```


not null,
not null,
not null




```
create table actor
(
    actor_id integer not null primary key,
    first_name varchar(45) not null,
    last_name varchar(45) not null,
    last_update timestamp not null
);
```

Performance Testing

Will this understand us and give us good information?

 Who acts in it?

 To see the list of actors who appeared in Academy Dinosaur, you can run the following query:

last_name	first_name
Guiness	Penelope
Gable	
Tracy	
Peck	
Cage	
Temple	Mena
Nolte	Warren
Kilmer	Oprah
Dukakis	Rock
Keitel	Mary

```
SELECT actor.first_name, actor.last_name
FROM actor
JOIN film_actor ON actor.actor_id = film_actor.actor_id
JOIN film ON film_actor.film_id = film.film_id
WHERE film.title = 'Academy Dinosaur';
```



Performance Testing

Will this understand us and give us good information?



Show me all the movies that actors in this movie have acted in



Sure! Here's the query to retrieve all the movies that actors in Academy Dinosaur have acted in:

title	
Northwest Polish	<pre>SELECT DISTINCT film.title FROM film JOIN film_actor ON film.film_id = film_actor.film_id JOIN actor ON film_actor.actor_id = actor.actor_id WHERE actor.actor_id IN (SELECT actor.actor_id FROM actor JOIN film_actor ON actor.actor_id = film_actor.actor_id JOIN film ON film_actor.film_id = film.film_id WHERE film.title = 'Academy Dinosaur') AND film.title <> 'Academy Dinosaur' LIMIT 100;</pre>
Instinct Airport	
Egg Igby	
Gilmore Boiled	
Seabiscuit Punk	
Luke Mummy	
Beach Heartbreakers	
Jumanji Blade	
Fantastic T	



Performance Testing

Will this understand us and give us good information?



Which actor has been in the most Documentary films?



To find the actor who has been in the most Documentary films, you can run the following query:

documentary_films	actor_name
6	Dustin Tautou



Which two actors have been in the most films together?



To find the two actors who have

actor_pair
Julia Mcqueen and Henry Be

```
SELECT CONCAT_WS(' and ', a1.first_name || ' ' ||  
a1.last_name, a2.first_name || ' ' || a2.last_name) AS  
actor_pair, COUNT(*) AS num_films_together  
FROM film_actor fa1  
JOIN film_actor fa2 ON fa1.film_id = fa2.film_id AND  
fa1.actor_id < fa2.actor_id  
JOIN actor a1 ON fa1.actor_id = a1.actor_id  
JOIN actor a2 ON fa2.actor_id = a2.actor_id  
GROUP BY a1.actor_id, a2.actor_id  
ORDER BY num_films_together DESC  
LIMIT 1;
```



Performance Testing

Will this understand us and give us good information?



Just tell me the synopsis of the movie



Sure! Here's the synopsis of the film Academy Dinosaur:

In this action-packed adventure, a group of paleontologists uncover a lost world filled with prehistoric creatures, including the fearsome Tyrannosaurus Rex. As they struggle to survive in this dangerous new environment, they must confront not only the dangers of the past, but also the moral quandaries of the present. With breathtaking special effects and a heart-stopping score, Academy Dinosaur is sure to thrill audiences of all ages.



Just tell me the synopsis (in English and Chinese)



The synopsis of 'Academy Dinosaur' is not available in multiple languages in the database. However, here is an SQL query that retrieves the synopsis in English:

description
A Epic Drama of a Feminist And a Mad Scientist who must Battle a Teacher in The Canadian Rockies

Note that the result will be in English only.

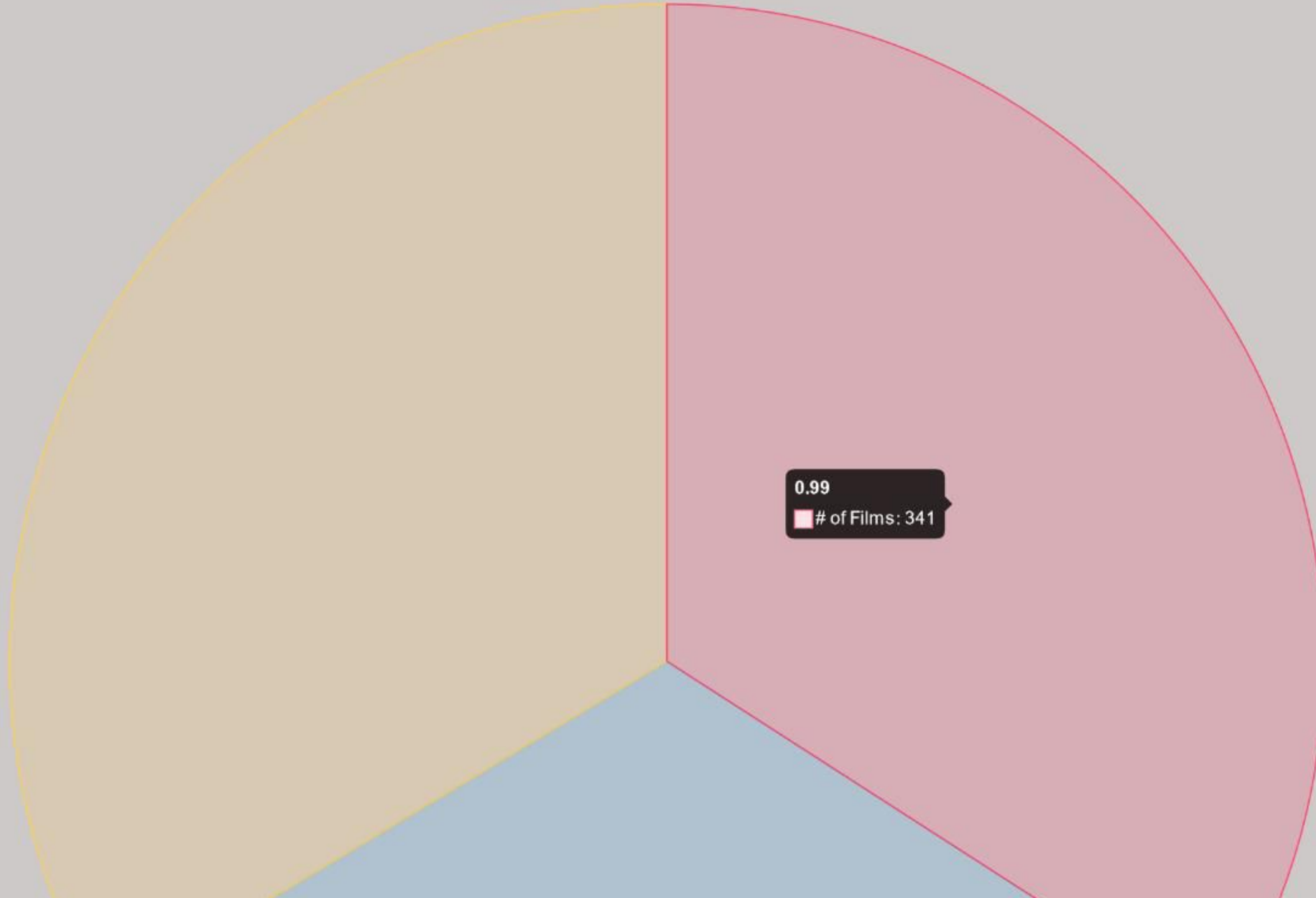


Show a pie chart showing number of films in each rental rate



Number of Films in Each Rental Rate

0.99 2.99 4.99



What about Javascript?

Generating code to show the results in a chart

`conversation.messages.push({role: "system", content: `You are a query engine for a database called Film DB. The user will ask you questions. You may answer in free text or by running a query in the DB. Any sql you write will be executed and the user will see the results of that query in a DB table. Please always add LIMIT 20. Below is the DB schema for the PostgreSQL 14 database.``

Note also that you can answer by generating charts if appropriate. In this case, you should generate SQL which will run and store the results as a javascript array in the `\`window.rows\`` variable, e.g. `select eim_name from apps` might result in this javascript object being available:

```
window.rows = [  
  { "title": "Film 1" }, { "title": "Film 2" }  
]
```

If generating a chart:

- * You must use chart.js version 3 to generate a chart
- * The canvas ID should be `\`myChart\``
- * You must use data from the `\`window.rows\`` global variable
- * You may use moment.js for time and date manipulations
- * You should return the javascript in a markdown code block with the `\`javascript\`` language tag
- * You should return the SQL in a markdown code block with the `\`sql\`` language tag
- * Do not include any HTML tags - just the javascript

```
create table actor  
(  
  actor_id integer not null primary key,  
  first_name varchar(45) not null,  
  last_name varchar(45) not null,  
  last_update timestamp not null  
);
```



Show bar chart showing number of films per genre

To create a bar chart showing the number of films per genre, we can use the `category` and `film_category` tables.

Here is the SQL query:

```
```sql
SELECT c.name AS category, COUNT(fc.film_id) AS num_films
FROM category c
LEFT JOIN film_category fc
 ON c.category_id = fc.category_id
GROUP BY c.name
ORDER BY num_films DESC
LIMIT 20;
```
```

And here is the javascript code to generate the chart:

```
```javascript
// Get the data from the `window.rows` global variable
const data = window.rows.map(row => ({
 category: row.category,
 num_films: row.num_films
}));

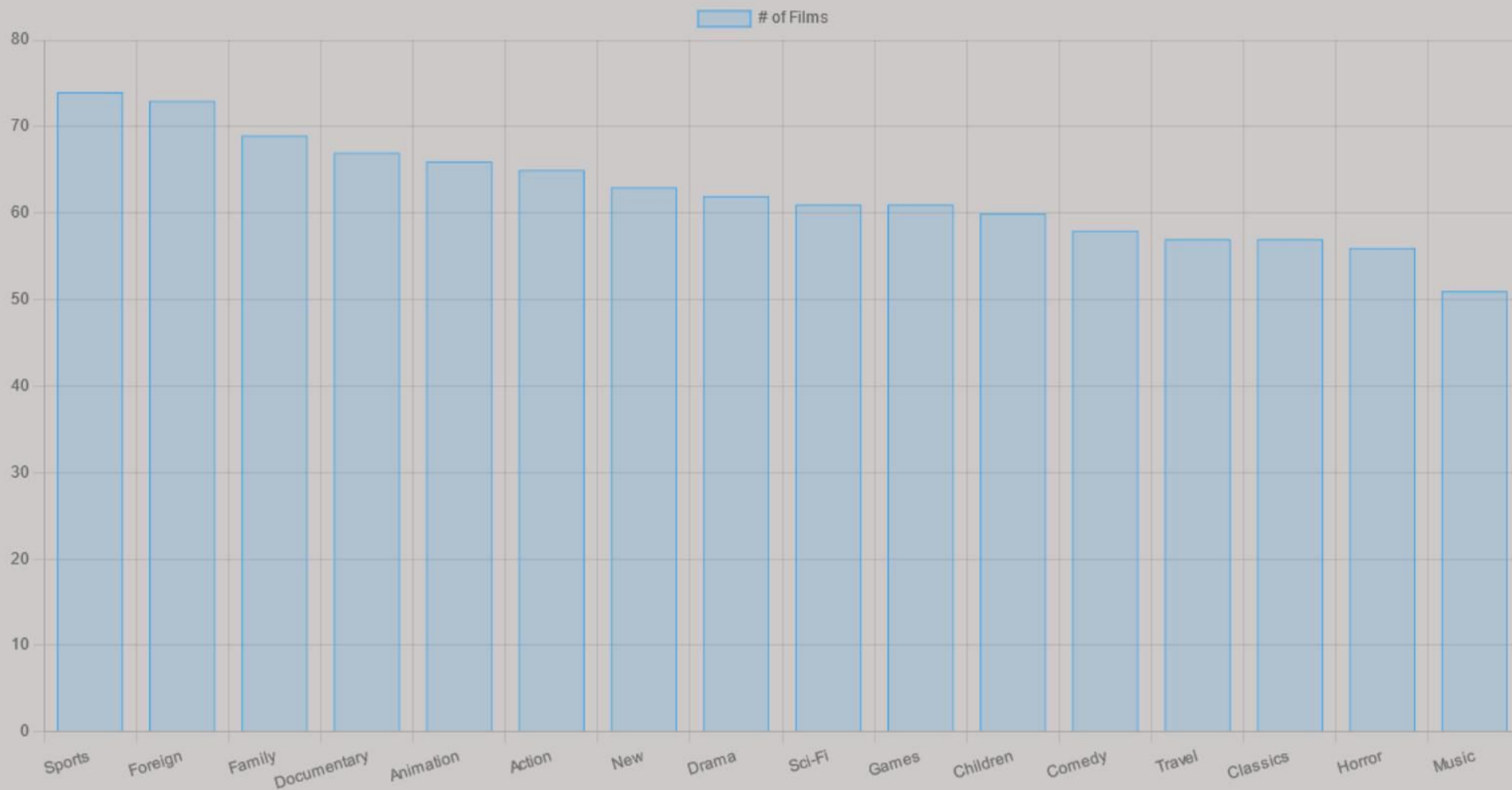
// Extract the categories and number of films from the data
const labels = data.map(row => row.category);
const values = data.map(row => row.num_films);

// Set up the chart data
const chartData = {
 labels: labels,
 datasets: [{
```





Show bar chart showing number of films per genre



# Learnings

- The prompt is very important – experiment until it gives behaviour you need
- Be explicit and demanding: e.g. say exactly which version of a library to use
  - Show it example output you want it to generate, e.g. show it some JSON
    - Treat the generated code as if it is unsafe – execute in a sandbox
- It will make things up or be over-confident – implement error detection and retries

## Conclusion

This is a new super power

We can do things now that a year ago were unthinkable

But it is difficult to get a language model to do what you want!

Handling security and incorrect responses is tricky.

Coding against a model is a skill that needs to be practiced.

So time to get experimenting!

Thank you

