NAF
NETWORK AUTOMATION FORUM

{networkphil}

Using Large Language Models in Network Operations

The Human Factor

AUTOCON 2

THE NETWORK AUTOMATION CONFERENCE

What do we get out of it?

- **Paul**, a Senior Engineer for a national healthcare org
- Recently discovered Guinness chocolate cake and makes it every time our families get together
- Already has a solid operations workflow and is skeptical about new tech and marketing hype

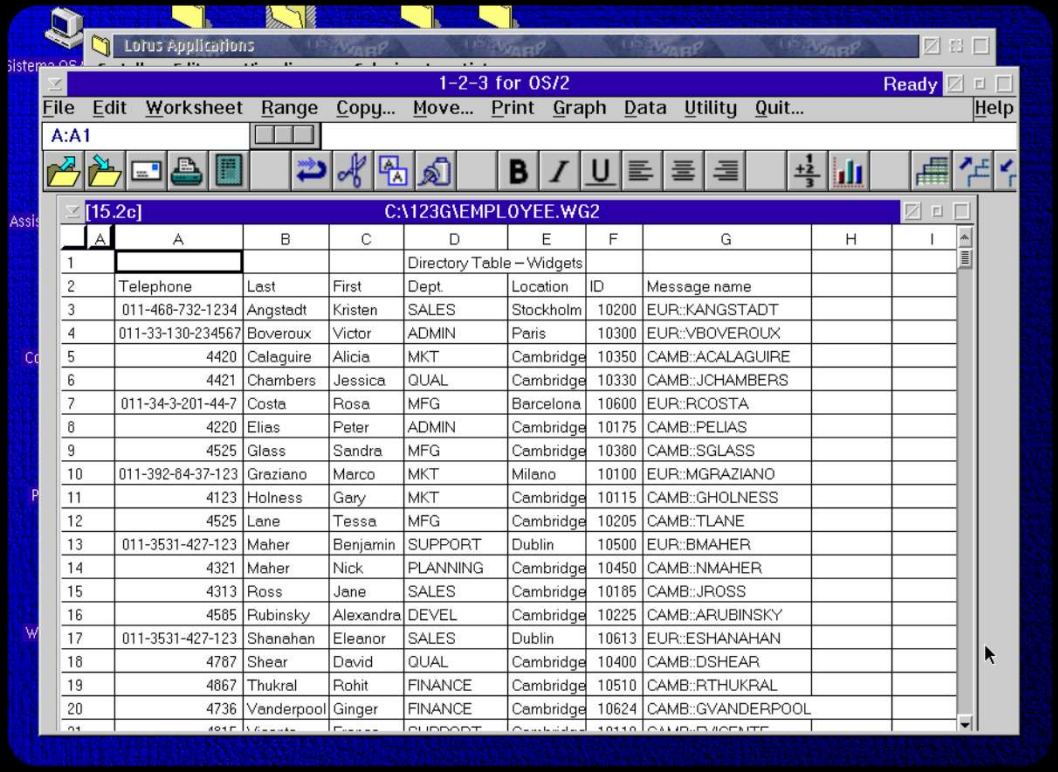
- **Hank**, a Master Plumber/GM at plumbing store chain
- Can't stop talking about taking his new camper across the country on an epic road trip
- Threatens to quit every day because he can't find anything in the inventory system and has lost customers because of botched orders

The human factor

- We’re augmenting a human being
- We’re making the work people do easier, faster, and more insightful
- We’re democratizing information among job roles, skill levels,teams, etc

The human factor

HP-45 Calculator



	A	B	C	D	E	F	G	H	I
1				Directory Table – Widgets					
2	Telephone	Last	First	Dept.	Location	ID	Message name		
3	011-468-732-1234	Angstadt	Kristen	SALES	Stockholm	10200	EUR::KANGSTADT		
4	011-33-130-234567	Boveroux	Victor	ADMIN	Paris	10300	EUR::VBOVEROUX		
5	4420	Calaguire	Alicia	MKT	Cambridge	10350	CAMB::ACALAGUIRE		
6	4421	Chambers	Jessica	QUAL	Cambridge	10330	CAMB::JCHAMBERS		
7	011-34-3-201-44-7	Costa	Rosa	MFG	Barcelona	10600	EUR::RCOSTA		
8	4220	Elias	Peter	ADMIN	Cambridge	10175	CAMB::PELIAS		
9	4525	Glass	Sandra	MFG	Cambridge	10380	CAMB::SGLASS		
10	011-392-84-37-123	Graziano	Marco	MKT	Milano	10100	EUR::MGRAZIANO		
11	4123	Holness	Gary	MKT	Cambridge	10115	CAMB::GHOLNESS		
12	4525	Lane	Tessa	MFG	Cambridge	10205	CAMB::TLANE		
13	011-3531-427-123	Maher	Benjamin	SUPPORT	Dublin	10500	EUR::BMAHER		
14	4321	Maher	Nick	PLANNING	Cambridge	10450	CAMB::NMAHER		
15	4313	Ross	Jane	SALES	Cambridge	10185	CAMB::JROSS		
16	4585	Rubinsky	Alexandra	DEVEL	Cambridge	10225	CAMB::ARUBINSKY		
17	011-3531-427-123	Shanahan	Eleanor	SALES	Dublin	10613	EUR::ESHANAHAN		
18	4787	Shear	David	QUAL	Cambridge	10400	CAMB::DSHEAR		
19	4867	Thukral	Rohit	FINANCE	Cambridge	10510	CAMB::RTHUKRAL		
20	4736	Vanderpool	Ginger	FINANCE	Cambridge	10624	CAMB::GVANDERPOOL		

Lotus 123


NAF
NETWORK AUTOMATION FORUM
What are LLMs, anyway?

What are LLMs?

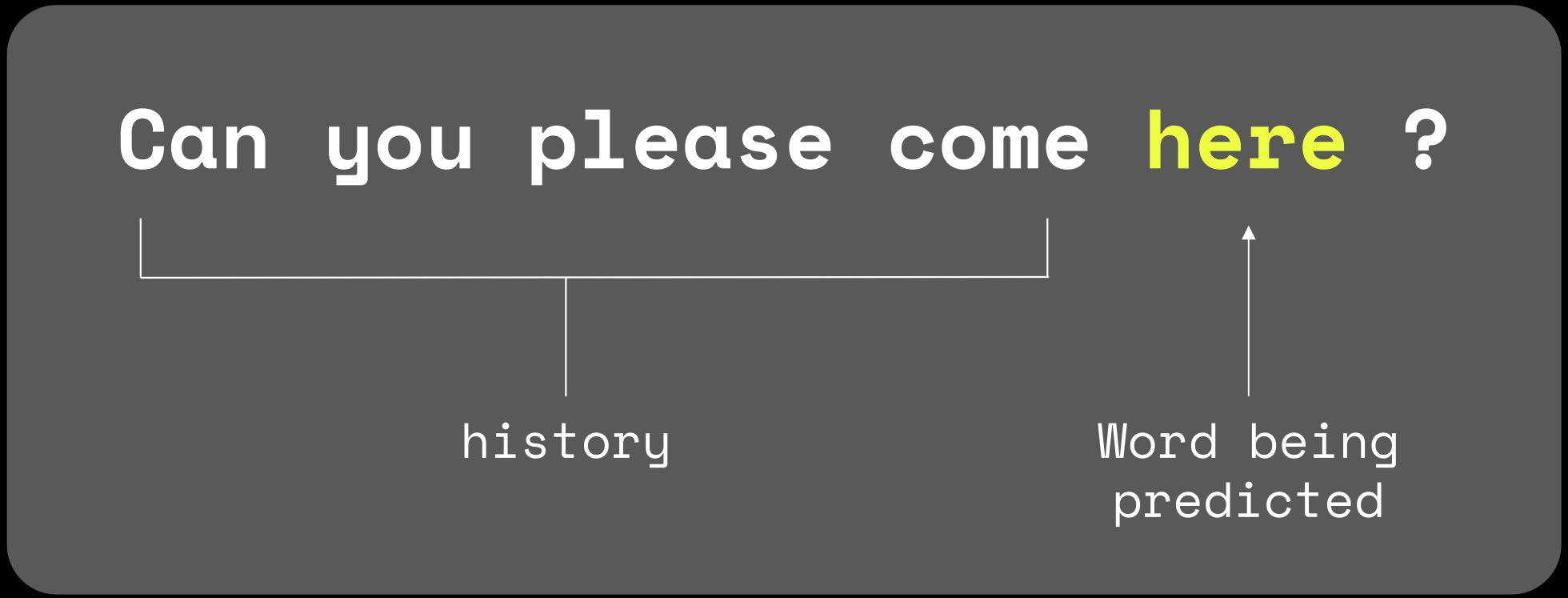
Probabilistic Models

- Mathematical models that predict the next word
- No inherent intelligence, human-like reasoning, or true thinking ability (some experts are starting to disagree)
- Not deterministic which is why they produce different outputs to the same prompt

How They Work

- Semantic framework built on neural network trained on a large body of text
- Represent words as embeddings in vector space so they can see how words relate to each other (semantic proximity)
- Sees the relationships among words and therefore can predict how to respond

It works, but we're missing context




JALAD. HIS NETWORK DOWN.

Transformers and Attention

- Modern LLMs are built on the Transformer model that captures longer range dependencies in text (context)

- Uses an attention mechanism that assigns weights to different data (words) to help understand context

Challenges

- Hallucinations
- Real-time telemetry data
- Handling diverse data types, sources, etc.
- Privacy, regulatory concerns
- Yet another thing to learn. . .

Human in the loop

Involve human oversight and/or intervention in the training, architecture, deployment, and real-time operations of LLMs.


NAF
NETWORK AUTOMATION FORUM
Use cases

Querying large datasets

LLMs can help anyone query large, diverse, and even real-time datasets very quickly and easily.

Scenario

You're an engineer working on a forensics analysis. You need to identify:

1. Application network traffic
2. Egressing a specific VPC in US-EAST-1
3. Over the last 24 hours
4. Destined for any embargoed countries

Programmatic data analysis

LLMs facilitate an advanced data analysis workflow that anyone can use.

Scenario

You're t-shooting a slow app hosted in Azure, but the app is slow for folks only in the Chicago branch and only at certain times of the day every day. You need to figure out:

1. What do people mean by "slow"?
2. What's unusual/different/weird about the app or the network?
3. Why only certain times of the day?
4. Why only Chicago?

Automated NetOps workflows

LLMs can facilitate a sophisticated autonomous network operations workflow.

What's the benefit?

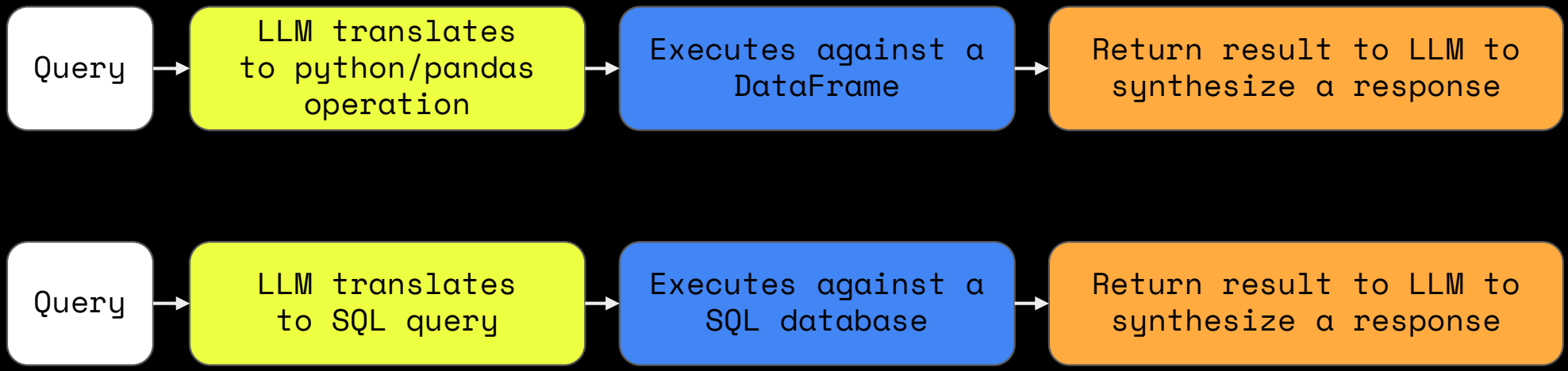
- Reduction of manual clue-chaining* and root cause analysis
- Get insight we couldn't otherwise see (due to lack of ability, skill, time, etc.)
- Potential for a more autonomous NetOps workflow


NAF
NETWORK AUTOMATION FORUM
How can we actually use LLMs in NetOps?

Add data and context to your prompt

```
>>> Analyse these logs and tell me how many authentication failures there are.
(log output omitted)

1. **Invalid Password:**
            * 2024-11-05T01:17:45Z - admin (192.168.101.11)
            * 2024-11-05T03:25:37Z - guest (192.168.103.13)
            * 2024-11-05T04:35:02Z - root (192.168.104.14)
            * 2024-11-05T17:45:40Z - admin (192.168.118.28)
            * 2024-11-06T00:45:10Z - user (192.168.124.34)
            * 2024-11-06T03:05:40Z - test (192.168.126.36)
            * 2024-11-06T05:25:10Z - test (192.168.128.38)
            Total number of Invalid Password failures: 7
2. **User Not Found:**
            * 2024-11-05T13:05:40Z - user (192.168.114.24)
            * 2024-11-05T15:25:10Z - test (192.168.116.26)
            * 2024-11-06T03:05:40Z - test (192.168.126.36)
            Total number of User Not Found failures: 3

3. **Invalid Password (Root):**
            * 2024-11-05T11:45:10Z - root (192.168.112.22)
            * 2024-11-06T04:15:25Z - admin (192.168.127.37) is not in the original list, it was an
error
            Total number of Invalid Password (Root) failures: 1

The total number of authentication failures is the sum of the individual failure counts:

7 (Invalid Password) + 3 (User Not Found) + 1 (Invalid Password (Root)) = 11
```

Use the LLM to generate your queries



Use the LLM to generate your queries

PG show me all application traffic egressing AWS US-EAST-1 for the last 24 hours destined for embargoed countries

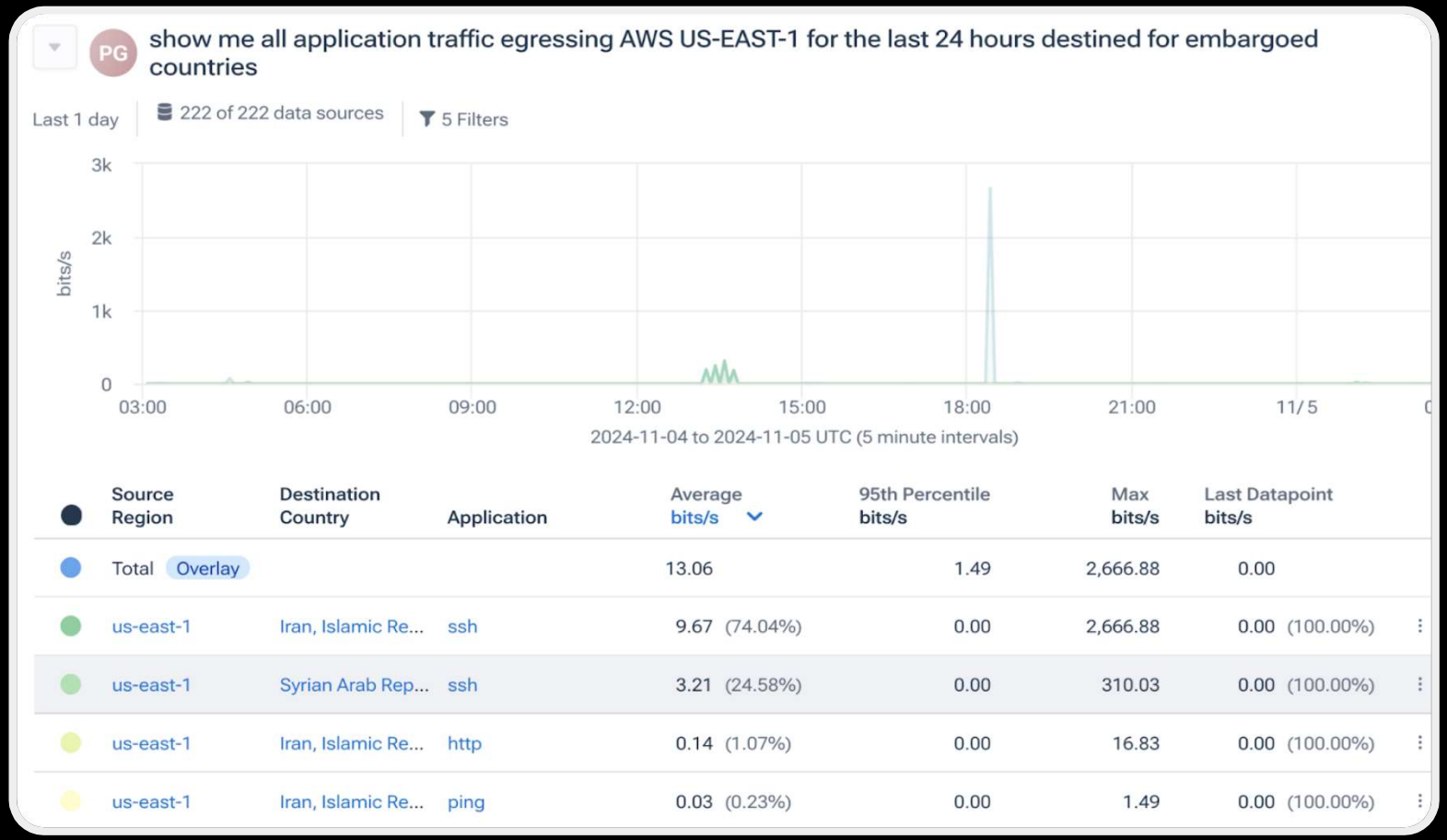
```sql
SELECT
  kt_aws_src_region,
  CASE WHEN kt_aws_src_region IS NULL
  OR kt_aws_src_region = '' THEN '---' ELSE kt_aws_src_region END AS kt_aws_src_region,
  dst_geo,
  CASE WHEN dst_geo IS NULL
  OR dst_geo = '' THEN '---' ELSE dst_geo END AS Geography_dst,
  application,
  CASE WHEN application IS NULL
  OR application = '' THEN '---' ELSE application END,
  array_agg(f_sum_both_bytes) as f_sum_both_bytes,
  array_agg(
    extract(
      epoch
      from
        i_start_time
    )
  ) as timestamps,
  array_agg(i_duration) as durations
FROM
  (
    SELECT
      i_start_time,
      kt_aws_src_region,
      dst_geo,
      application,
      sum(both_bytes) as f_sum_both_bytes,
      max(i_duration) as i_duration,
      row_number() OVER (
        PARTITION BY i_start_time
        ORDER BY
          sum(both_bytes) DESC
      )
    FROM
```

LLM understands our words and generates SQL which our workflow then applies to our db

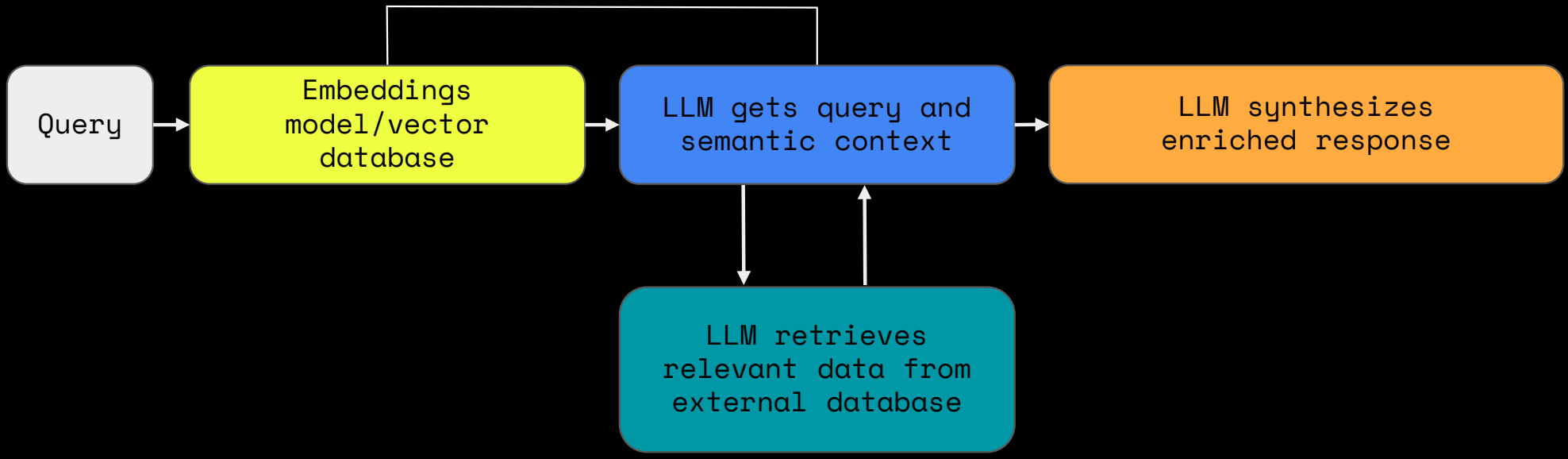
```json
{
  "version": 4,
  "queries": [
    {
      "bucket": "Left +Y Axis",
      "isOverlay": false,
      "query": {
        "all_devices": true,
        "aggregateTypes": [
          "avg_bits_per_sec",
          "p95th_bits_per_sec",
          "max_bits_per_sec"
        ],
        "aggregateThresholds": {},
        "bracketOptions": null,
        "hideCidr": false,
        "cidr": 32,
        "cidr6": 128,
        "customAsGroups": true,
        "cutFn": {},
        "cutFnRegex": {},
        "cutFnSelector": {},
        "depth": 75,
```

Query response returned to LLM which translates to JSON

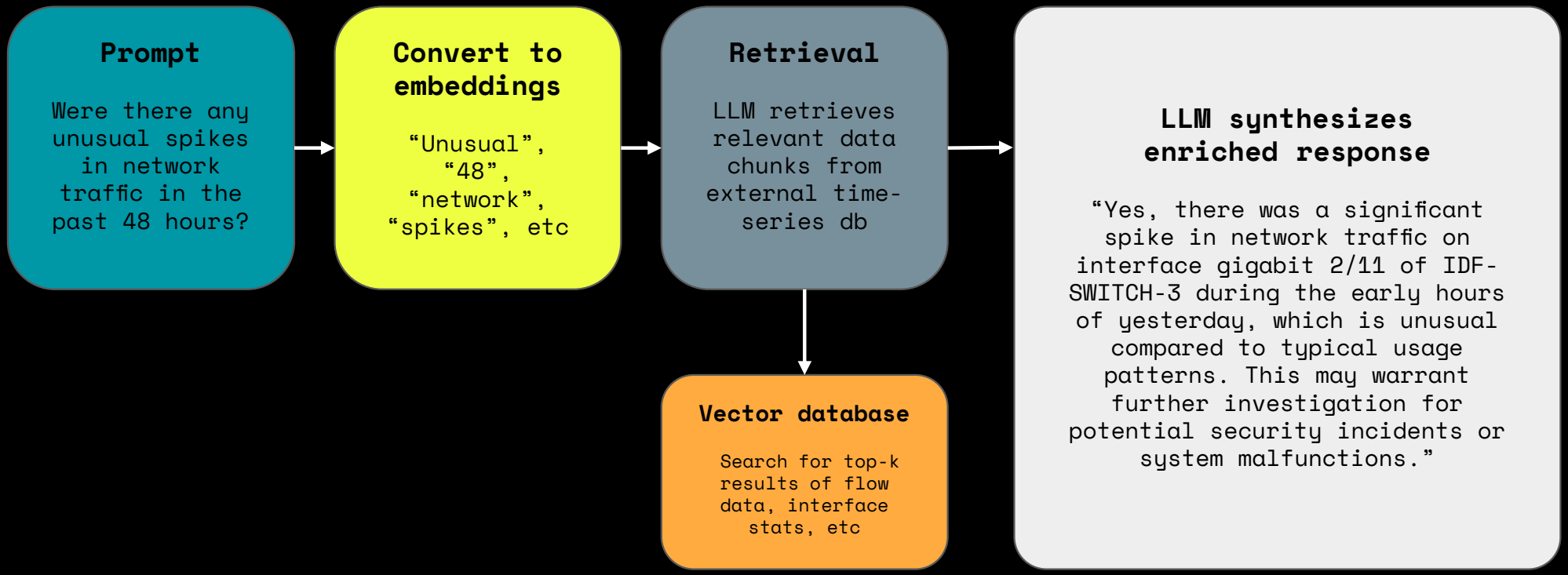
Use the LLM to generate your queries



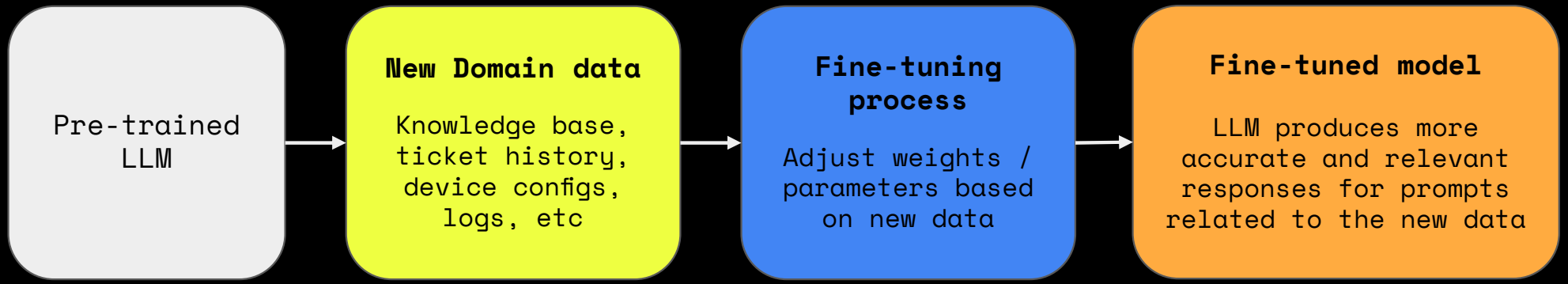
Use an LLM in a RAG system with a vector database



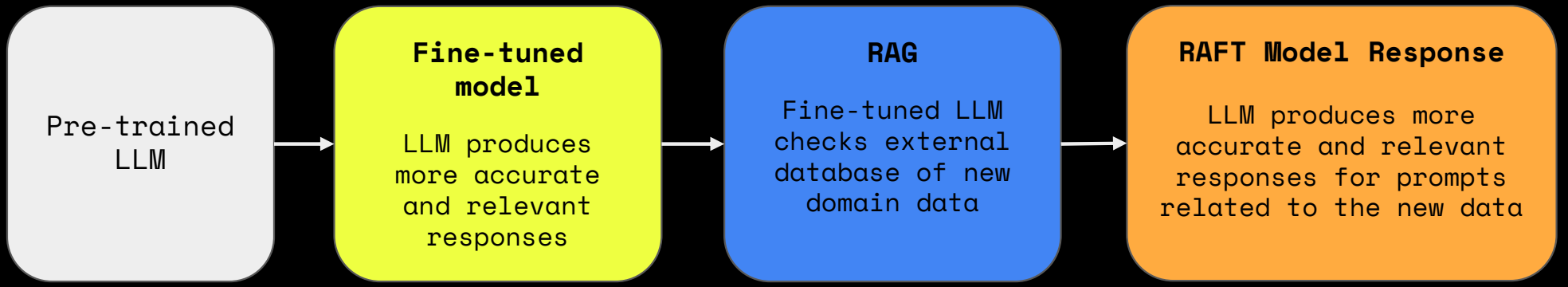
Use an LLM in a RAG system with a vector database



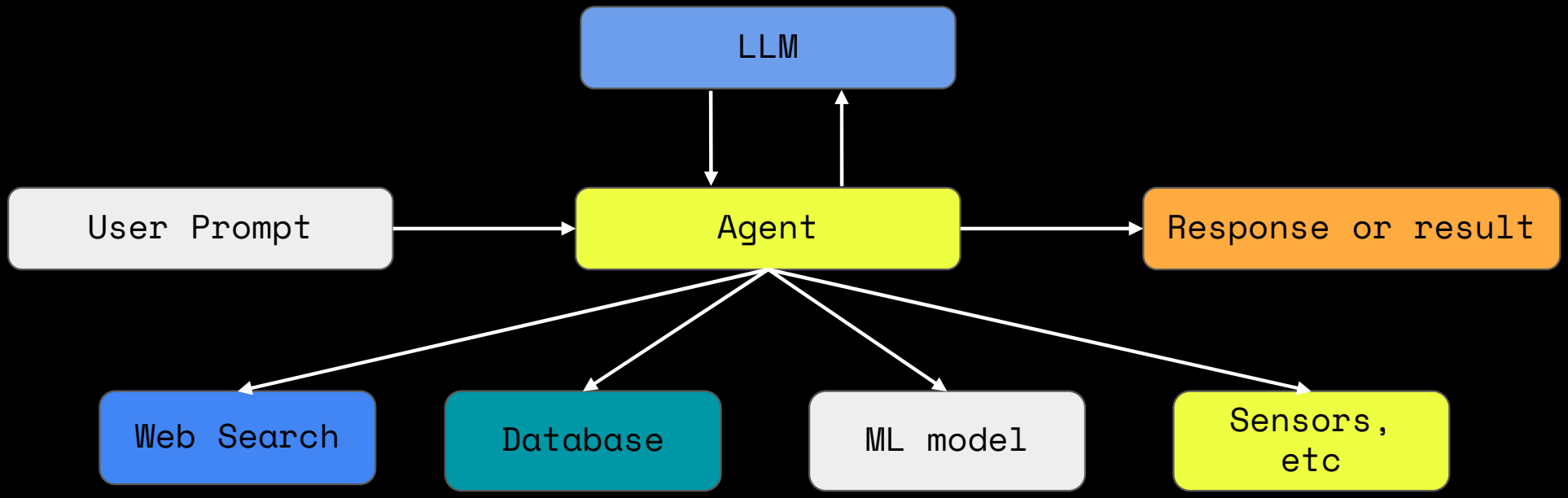
Fine-tune an LLM for your specific domain and needs



Combine fine-tuning and RAG (RAFT)



Using agents and LLMs together




NAF
NETWORK AUTOMATION FORUM
Getting started
(like right now)

How to get started right now

1. Download Ollama (do it right now)

1. Go to *huggingface.co* and download a smaller model like Llama 3.2 or whatever suits your fancy

1. Sign up for and/or download a free vector database like Chroma (local) or the free version of Pinecone (hosted) to use for RAG and get your sweet sweet network data in there

How to get started right now

4. Use a programmatic workflow framework like LangChain to tie the pieces together in a pretty pythonic bow

4. Use Streamlit to make a quick and easy web app and impress your friends

4. And of course go check out Amazon Bedrock or Azure AI Studio for and easy way to build an LLM workflow with dropdowns, menus, and wizards

So what about our new friends?

Paul, Sr. Engineer

Hank, Master Plumber


NAF
NETWORK AUTOMATION FORUM
Questions?