POSETTE: An Event for Postgres 2025

Hosted by Microsoft

Building Intelligent Applications with Graph-Based RAG on PostgreSQL

Abe Omorogbe

Microsoft, Senior PM

June 10-12, 2025

#PosetteConf

ABE OMOROGBE

- Product Manager at Postgres AI team at Microsoft.
- My previous experiences includes working Azure AI and Databricks

· @_aiabe
· linkedin.com/in/abeomor
· github.com/AbeOmor

Has ChatGPT or Copilot ever given you an insufficient answer?

What talk is Abe Omorogbe's giving at POSETTE?

As of now, there is no public information indicating that Abe Omorogbe is scheduled to give a talk at POSETTE: An Event for Postgres 2025. While he is listed as a member of the organizing team for the event , the official schedule and speaker lineup have not been released yet. posetteconf.com +7 posetteconf.com

The schedule for POSETTE 2025 is expected to be announced in March 2025, following the timeline from previous years . Once the schedule is published, it will be available on the POSETTE website, where you can find details about all the sessions and speakers. posetteconf.com posetteconf.com +7

If you're interested in updates about the event, including speaker announcements and session details, you can subscribe to notifications on the POSETTE website.

Sources

BUSINESS INSIDER

Subscribe

DOW JONES +0.01% NASDAQ +0.53% S&P 500 +0.16% AAPL +1.27% NVDA -0.53% MSFT +0.34% AMZN +1.49% META +1.2% TSLA +2.9%

AI

AI chatbots spew out nonsense too often. But there's a solution: retrieval-augmented generation.

By Matthew S. Smith

Fortune + Follow 128.4K Followers

AI search engines are confidently wrong more than half the time when they cite sources, study finds

Over 60% of AI chatbot responses are wrong, study finds

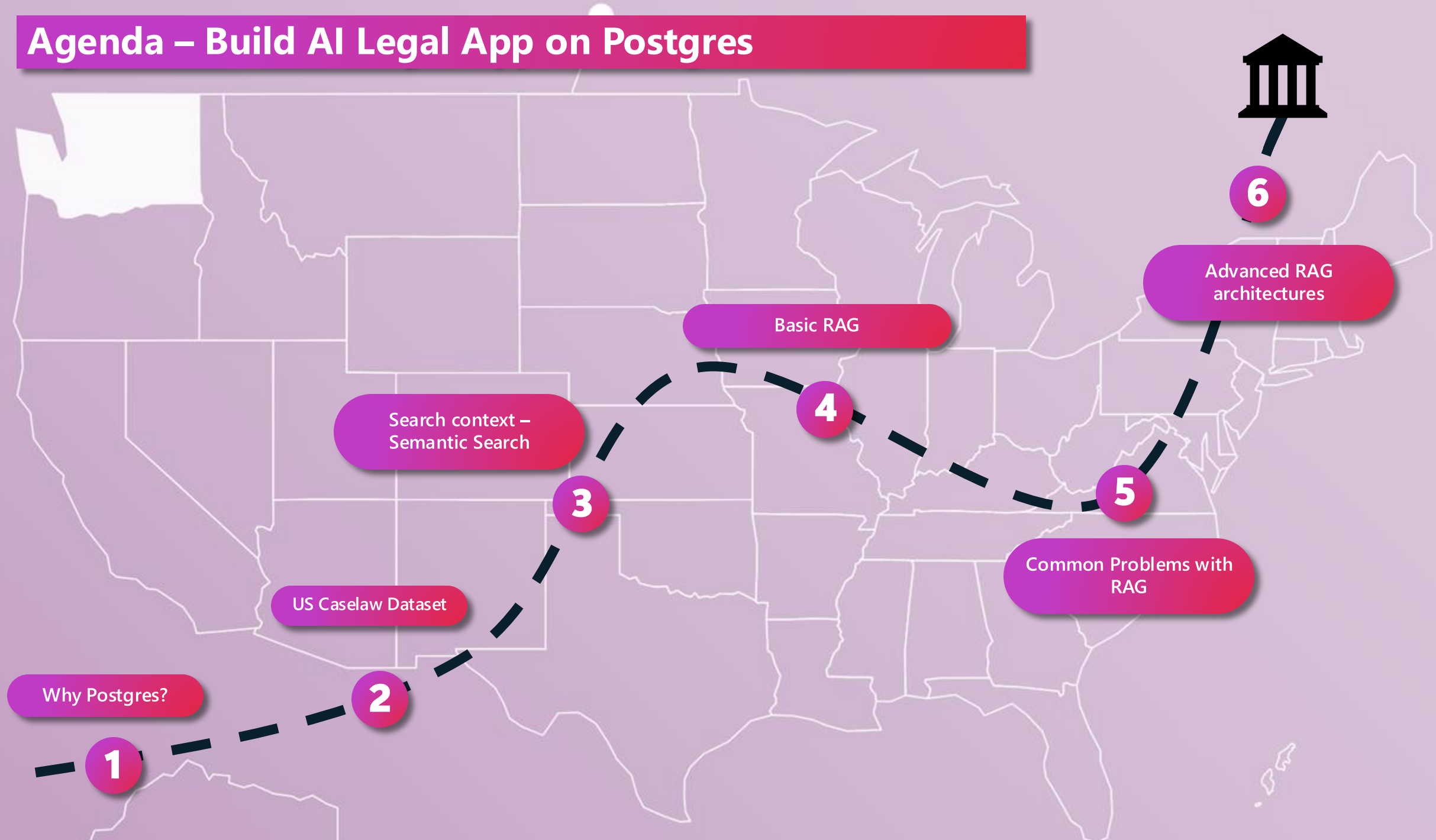
CNET + Follow 91.6K Followers

Gen AI's Accuracy Problems Aren't Going Away Anytime Soon, Researchers Say

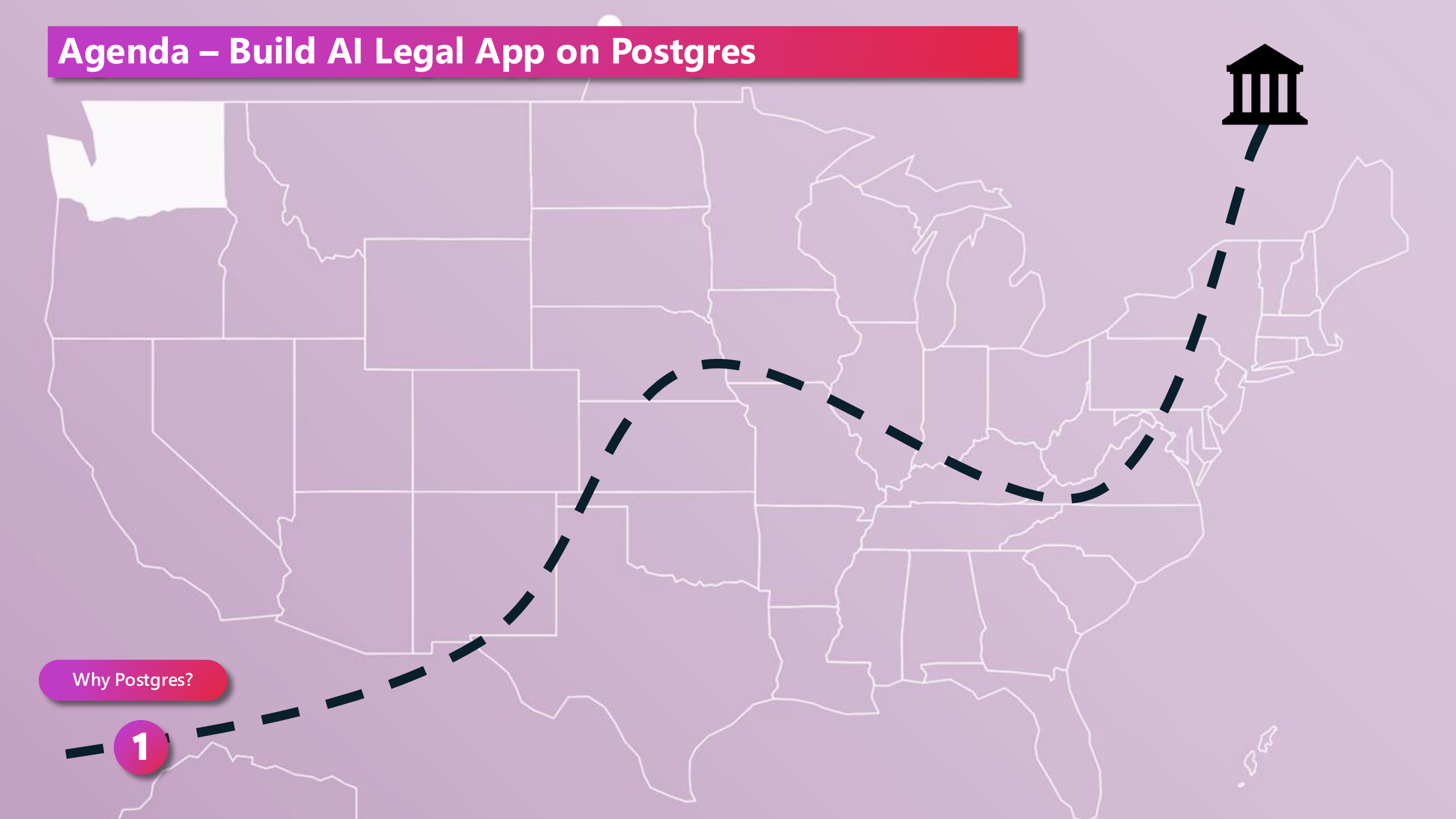
DIGITAL INFORMATION WORLD Advertise Web Tools

AI Hallucinations: Why Your Chatbot Might Be Lying to You

Agenda – Build AI Legal App on Postgres



Agenda – Build AI Legal App on Postgres



Why Postgres?

1

Benefits of Postgres for your GenAI app?

Postgres is the **#1 developer database** according to StackOverflow

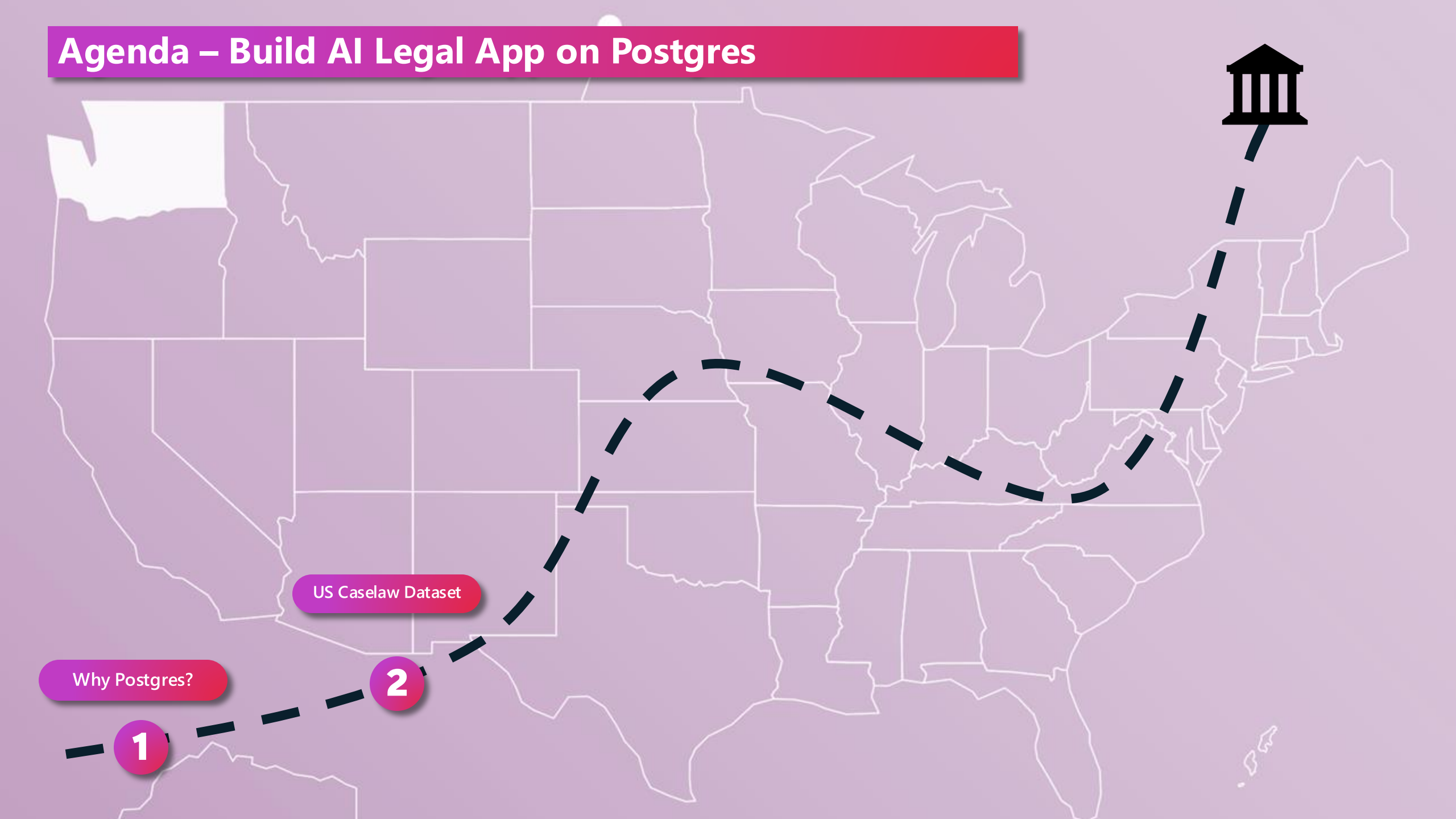
PostgreSQL offer features critical in building an AI app:

- AI model integration (azure_ai)
- Vector Search Capabilities (pgvector)
- Unstructured data support (JSONB)
- Graph Database (AGE)
- Geospatial Data support
- Extensive OSS frameworks
- Vibrant community and support

Legal Research Copilot app

- You're a legal researcher or lawyer focusing on cases in Washington state.
- You need the app to be able to find relevant previous cases to support your tenant rights and property dispute arguments.

Agenda – Build AI Legal App on Postgres



Caselaw Dataset (Washington State)

106,177

unique cases

7

legal reporters

581,403

pages scanned

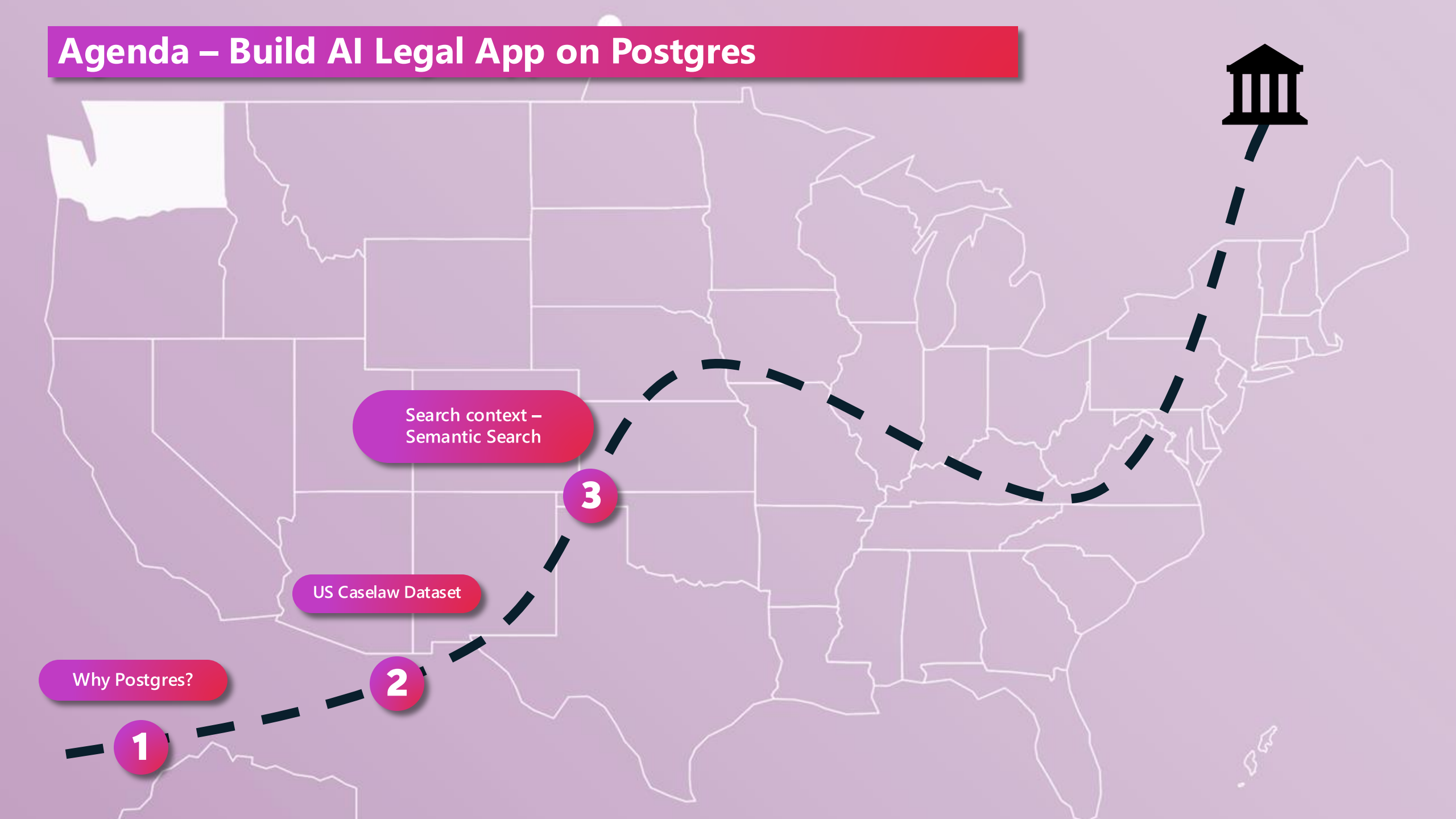
https://case.law/caselaw/#washington

The Data


```
[
  {
    "id": 5295672,
    "name": "George P. Hurd, Appellant, v. Henry Brisner and R. L. Hawthorne, Respondents",
    "name_abbreviation": "Hurd v. Brisner",
    "decision_date": "1891-10-14",
    "docket_number": "No. 207",
    "first_page": "1",
    "last_page": "6",
    "citations": [
      {
        "type": "official",
        "cite": "3 Wash. 1"
      }
    ],
    "court": {
      "name_abbreviation": "Wash.",
      "id": 9029,
      "name": "Washington Supreme Court"
    },
    "jurisdiction": {
      "id": 38,
      "name_long": "Washington",
      "name": "Wash."
    },
    "cites_to": [
      {
        "cite": "9 Kan. 632",
        "category": "reporters:state",
        "reporter": "Kan.",
        "case_ids": [91904],
        "opinion_index": -1,
        "case_paths": [
          "/kan/9/0632-01"
        ]
      },
      {
        "cite": "6 Col. 314",
        "category": "reporters:state",
        "reporter": "Col.",
        "opinion_index": -1
      },
      {
        "cite": "47 Iowa, 236",
        "category": "reporters:state",
        "reporter": "Iowa",
        "case_ids": [2331774],
        "opinion_index": -1,
        "case_paths": [
          "/iowa/47/0236-01"
        ]
      },
      {
        "cite": "34 La. Ann. 407",
        "category": "reporters:state",
        "reporter": "La. Ann.",
        "opinion_index": -1
      },
      {
        "cite": "70 Tex. 139",
        "category": "reporters:state",
        "reporter": "Tex.",
        "case_ids": [2194204],
        "opinion_index": -1,
```


name text	opinion text
Le Vette v. Hardman Estate	"Morris, J.\nAppeal from an order of nonsuit and dismissal, in an action brought by a tenant to recover dam
Wilkening v. Watkins Distributors, Inc.	"[As amended by order of the Court of Appeals October 19, 1989.]\nMunson, J.\nWatkins Distributors, Inc.,
Stevens v. King County	"Donworth, J.\nRespondents, as plaintiffs, brought this action against both the city of Seattle and Kang cou
Woolworth Co. v. City of Seattle	"Mitchell, J.\nAppellant, plaintiff below, was engaged in the mercantile business in the Arcade Building, in S
Uhl Bros. v. Hull	"Holcomb, J.\nAppellant, subtenant of a former lessee of a former owner of certain premises in Seattle, su
Tombari v. City of Spokane	"Millard, J.\nThis action was brought to recover for damages alleged to have been sustained as a result of
Imperial Candy Co. v. City of Seattle	"Main, J.\nThe purpose of this action was to recover damages to personal property caused by the breaking
Frisken v. Art Strand Floor Coverings, Inc.	"Rosellini, J.\nThe respondent, Florence Frisken, is the owner of a building in Shelton, Washington, occupie
Downie v. City of Renton	"Millaed, J.\n-This action was brought to enjoin the city of Renton from discharging waste water from its re
United Mutual Savings Bank v. Riebli	"Ott, J.\nOctober 30,1954, the Washington Building Company leased "that certain store space known as 10
Cordes v. Guy Investment Co.	"Parker, J.\nThe plaintiff, Cordes, seeks recovery of damages, claimed as the result of the failure of his land
Martindale Clothing Co. v. Spokane & Easter...	"Parker, J.\nThis action was commenced by the plaintiff to recover damages claimed to have resulted to it
Miller v. Vance Lumber Co.	"Parker, J.\nThe plaintiffs, Miller and wife, commenced this action in the superior court for King county see
DeHoney v. Gjarde	"Fullerton, J.\nOn April 5, 1922, J. H. DeHoney and wife entered into a contract with Peder P. Gjarde, by the
City of Spokane v. Fisher	"Tolman, J.\nThe city of Spokane was made defendant in a suit for personal injuries, brought by one Nicoli
Wood v. City of Tacoma	"Ellis, J.\nThis is an appeal from a judgment of non-suit and dismissal of an action to recover damages co
Wilber Development Corp. v. Les Rowland C...	"Rosellini, J.\nThis is an inverse condemnation action which the trial court dismissed on motions of the de
Wolten Grocery Co. v. Puget Sound Bridge & ...	"Beals, J.\n— Plaintiff corporation is the owner of a lot in the city of Port.Angeles, lying near the margin of t
Swanson v. White & Bollard, Inc.	"Geraghty, J.\nThis is an appeal from a judgment of dismissal entered after the court had sustained a chall

Agenda – Build AI Legal App on Postgres



How do we use this data to start building this Legal Copilot App

Search Pattern - ILIKE

SELECT id, name, opinion Untitled-2

```
1   SELECT id, name, opinion
2   FROM cases
3   WHERE opinion ILIKE '%Water leaking into the apartment from the floor above';
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS AZURE POSTGRESQL QUERY RESULTS

Results Messages

	id	name	opinion

Search Phrase – FULL TEXT

```sql
ALTER TABLE cases
ADD COLUMN textsearch tsvector
GENERATED ALWAYS AS (to_tsvector('english', name || LEFT(opinion, 8000))) STORED;

SELECT id, name, opinion
FROM cases
WHERE textsearch @@ websearch_to_tsquery('Water leaking into the apartment from the floor above.');
```

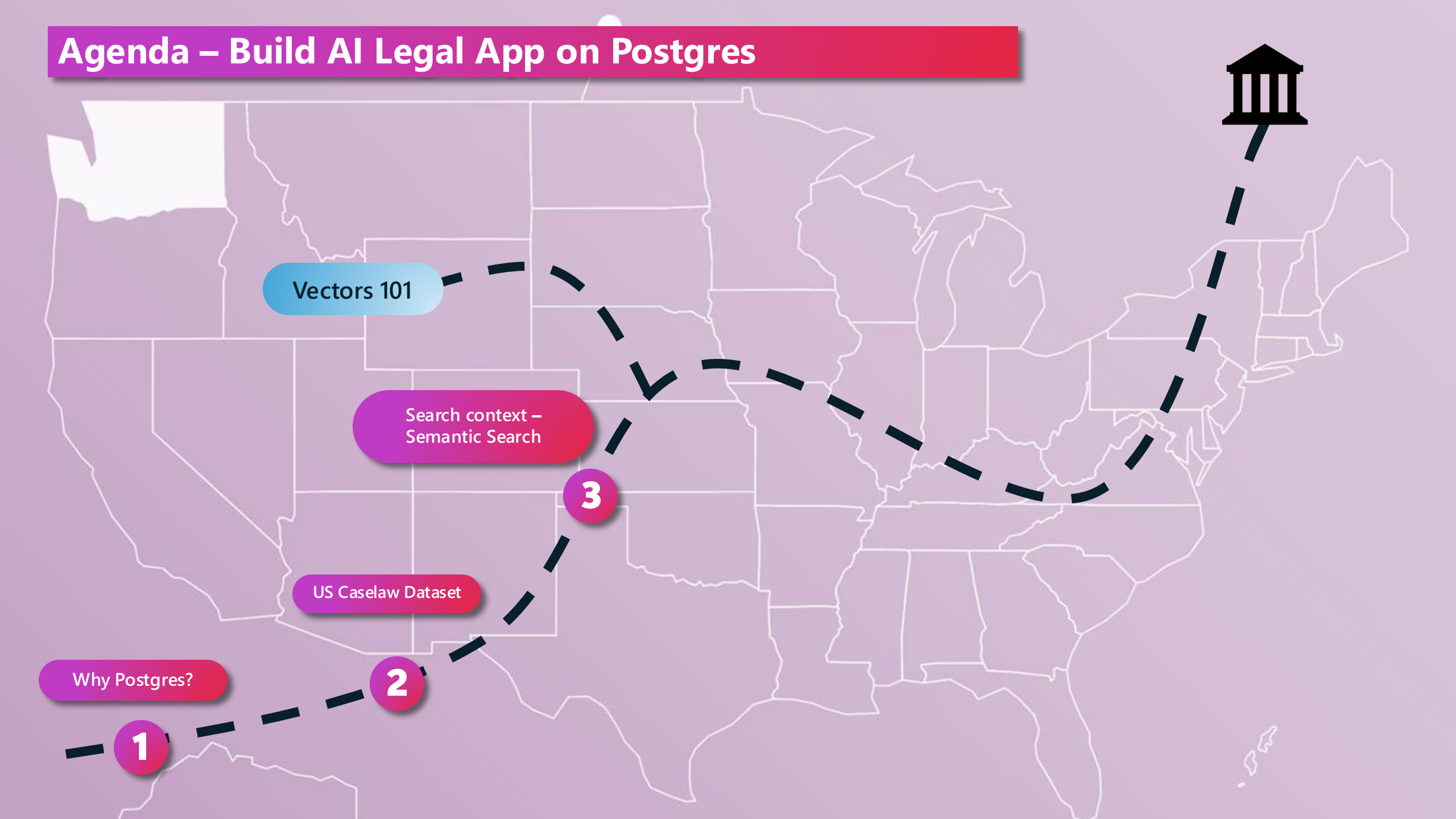
PROBLEMS 3 | OUTPUT | DEBUG CONSOLE | TERMINAL | PORTS | AZURE | POSTGRESQL QUERY RESULTS | SPELL CHECKER 3

Results | Messages

id	name	opinion
4237124	Pham v. Corbett	"Spearman, C.J.\n¶1 Landlord Lang Pham brought this u…

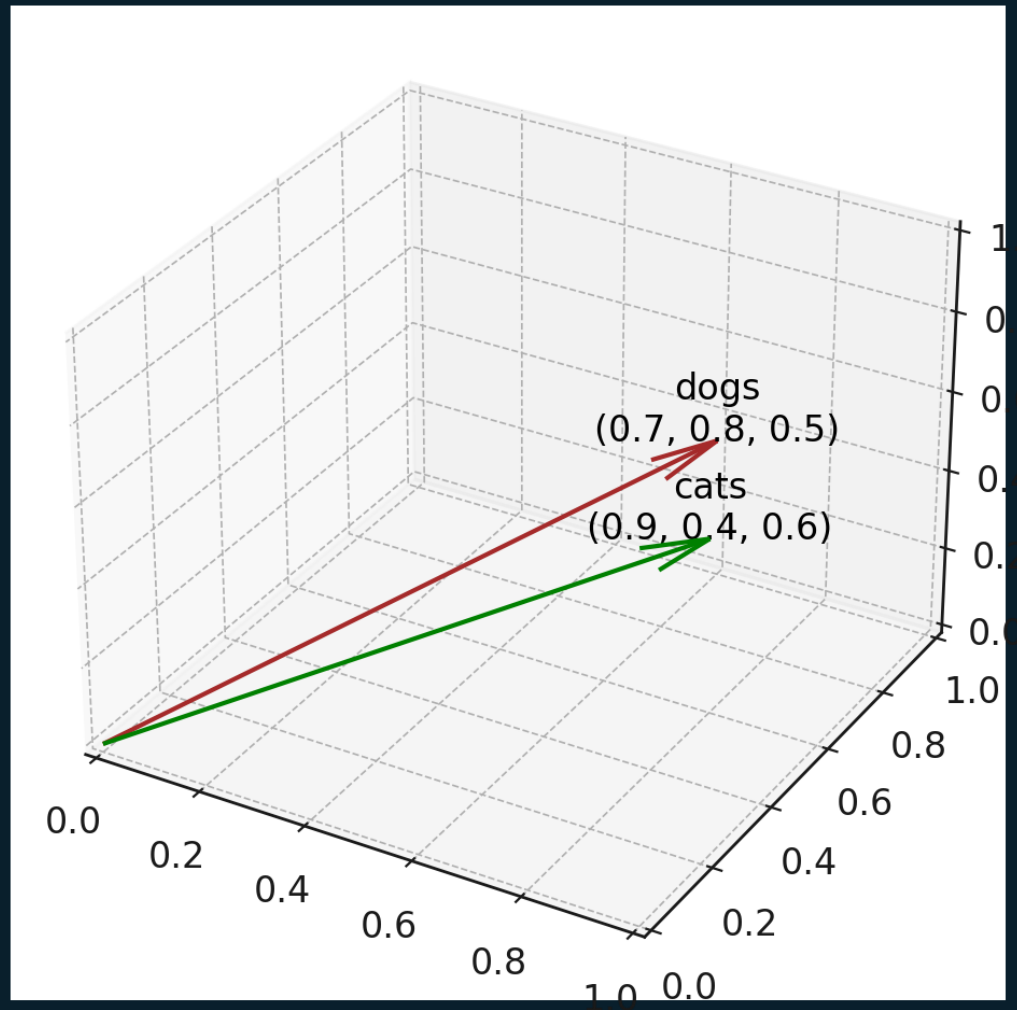
So how do we solve this?

Agenda – Build AI Legal App on Postgres



Vector 101

- Lists of numbers that represent items in a high-dimensional space.
- For example, a vector representing the string “**dogs**" might be [0.7, 0.8, 0.5].
- Each number in the vector is a dimension of the space.



Generating vectors

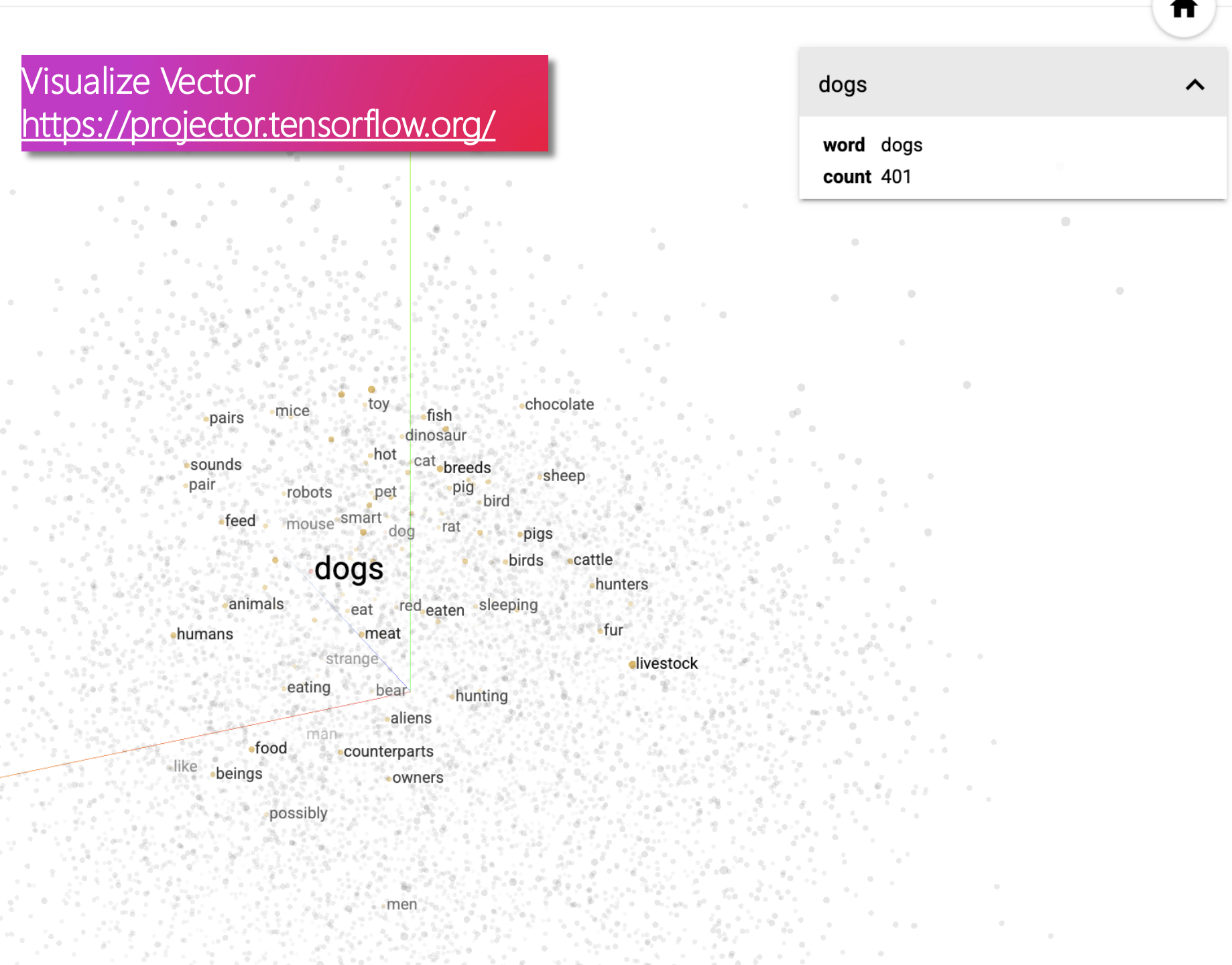
Use a model to generate vectors for items:

Input	→	Model	→	Vector
"dog"		word2vec		[0.017198, -0.007493, -0.057982, ..]
"cat"		word2vec		[0.004059, 0.06719, -0.093874, ...]

Model	Input types	Dimensions
Word2Vec	Word	50-300
OpenAI text-embedding-3	Text	256-3072
Cohere embed-v4.0	Text	256-1536
Azure Computer Vision Multi-modal	Text or Image	1024

Popular models (find more on HuggingFace)

Visualize Vector
https://projector.tensorflow.org/



dogs

word dogs
count 401

Search by
dogs .* word

neighbors 100

distance COSINE EUCLIDEAN

Nearest points in the original space:

dog	0.424
cats	0.496
animals	0.600
mice	0.600
cat	0.604
pet	0.607
breed	0.610
breeds	0.627
horses	0.634
cow	0.644
hunting	0.646
humans	0.650
fish	0.668
pigs	0.670
toy	0.675

Storing vectors in Postgres table

	name	opinions_vector
1	Le Vette v. Hardman Estate	[-0.0016331548,0.061744377,0.009570962,0.035774387,-0.0054566152,0.025488522…
2	Uhl Bros. v. Hull	[0.011663684,0.07727938,0.013002017,0.03337608,0.012336588,-0.0059552076,-0.…
3	Woolworth Co. v. City of Seattle	[0.010754321,0.033327676,0.0031798491,0.037272118,0.005627261,3.528203e-05,-…
4	Stevens v. King County	[0.003271023,0.02641315,0.052163452,0.04789815,-0.012983223,-0.0027828913,-0…
5	Wilkening v. Watkins Distributors, Inc.	[-0.0191751,0.074948214,0.028303407,0.035043657,-0.036456708,0.025124041,0.0…
6	Tombari v. City of Spokane	[0.013072358,0.0324271,0.008165626,0.063088655,0.0021425574,0.0009784038,-0.…
7	Downie v. City of Renton	[0.031188406,0.055722144,0.019521315,0.056150556,-0.008168392,-0.010131948,-…
8	Frisken v. Art Strand Floor Coverings, Inc.	[0.0037472202,0.041518908,-0.0063951095,0.04394401,0.011329315,-0.0045397608…
9	Imperial Candy Co. v. City of Seattle	[0.017007183,0.03166547,0.022344228,0.04451006,-0.017914034,0.01880602,-0.03…
10	City of Spokane v. Fisher	[0.0014129812,0.03918049,0.00015752178,0.038623422,0.024062268,0.008920834,-…

```
-- Add Embeddings
ALTER TABLE cases ADD COLUMN opinions_vector vector(1536);
UPDATE cases
SET opinions_vector = azure_openai.create_embeddings('text-embedding-3-small',
name || opinion, max_attempts => 5, retry_delay_ms => 500)::vector
WHERE opinions_vector IS NULL;
```

Storing vectors in Postgres table

	name	opinions_vector
1	Le Vette v. Hardman Estate	[-0.0016331548,0.061744377,0.009570962,0.035774387,-0.0054566152,0.025488522…
2	Uhl Bros. v. Hull	[0.011663684,0.07727938,0.013002017,0.03337608,0.012336588,-0.0059552076,-0.…
3	Woolworth Co. v. City of Seattle	[0.010754321,0.033327676,0.0031798491,0.037272118,0.005627261,3.528203e-05,-…
4	Stevens v. King County	[0.003271023,0.02641315,0.052163452,0.04789815,-0.012983223,-0.0027828913,-0…
5	Wilkening v. Watkins Distributors, Inc.	[-0.0191751,0.074948214,0.028303407,0.035043657,-0.036456708,0.025124041,0.0…
6	Tombari v. City of Spokane	[0.013072358,0.0324271,0.008165626,0.063088655,0.0021425574,0.0009784038,-0.…
7	Downie v. City of Renton	[0.031188406,0.055722144,0.019521315,0.056150556,-0.008168392,-0.010131948,-…
8	Frisken v. Art Strand Floor Coverings, Inc.	[0.0037472202,0.041518908,-0.0063951095,0.04394401,0.011329315,-0.0045397608…
9	Imperial Candy Co. v. City of Seattle	[0.017007183,0.03166547,0.022344228,0.04451006,-0.017914034,0.01880602,-0.03…
10	City of Spokane v. Fisher	[0.0014129812,0.03918049,0.00015752178,0.038623422,0.024062268,0.008920834,-…

pgvector

```
-- Add Embeddings
ALTER TABLE cases ADD COLUMN opinions_vector vector(1536);
UPDATE cases
SET opinions_vector = azure_openai.create_embeddings('text-embedding-3-small',
name || opinion, max_attempts => 5, retry_delay_ms => 500)::vector
WHERE opinions_vector IS NULL;
```

azure_ai

Search Phrase – Vector Search

ALTER TABLE cases ADD COLUMN opinions_ve Untitled-2 ●

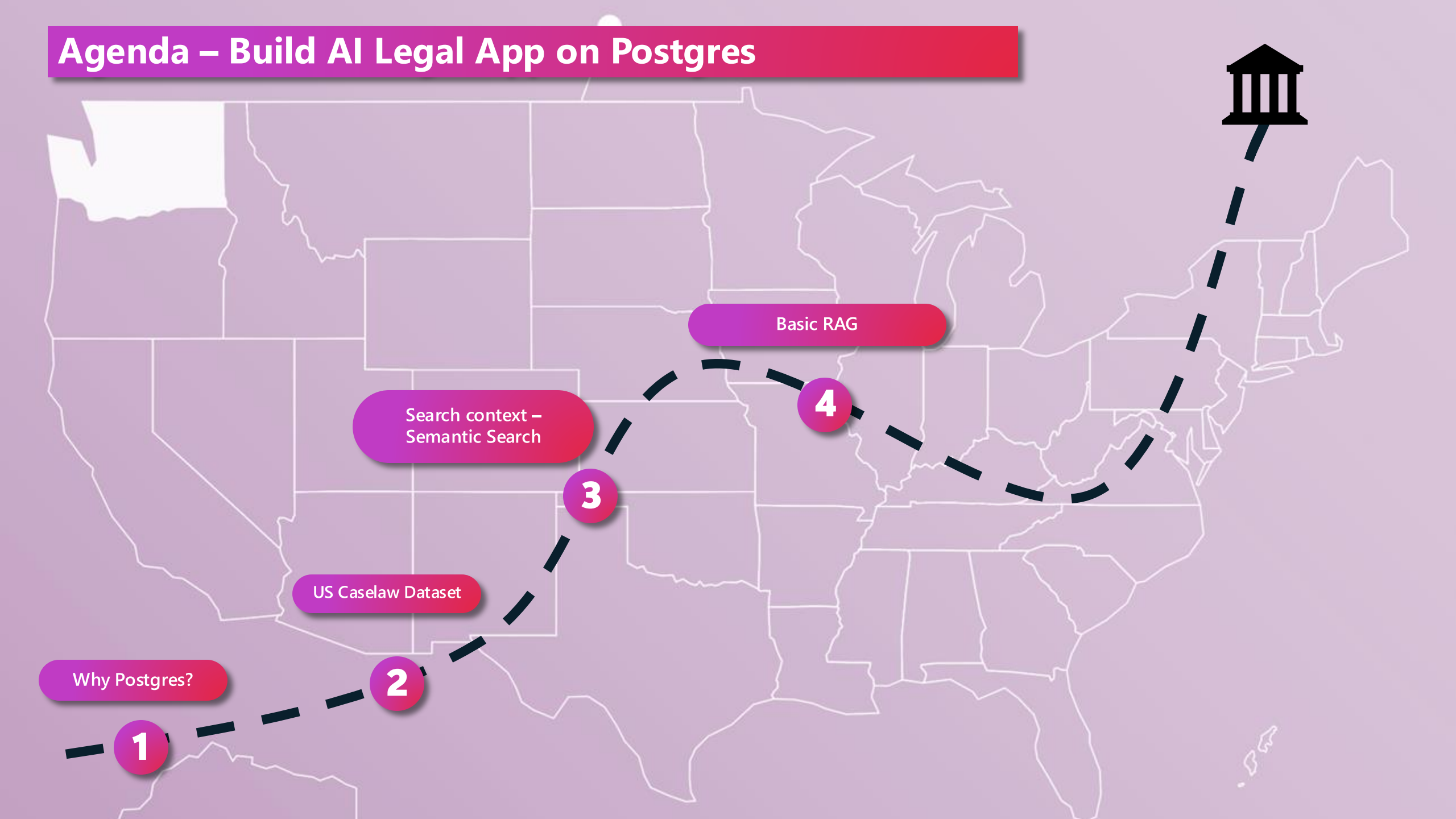
```
1  ALTER TABLE cases ADD COLUMN opinions_vector vector(1536);
2  UPDATE cases
3  SET opinions_vector = azure_openai.create_embeddings('text-embedding-3-small', name || opinion, max_attempts => 5, ret
4  WHERE opinions_vector IS NULL;
5
6  SELECT
7      id, name, opinion
8  FROM
9      cases
10 ORDER BY opinions_vector <=> azure_openai.create_embeddings('text-embedding-3-small', 'Water leaking into the apartmer
11 LIMIT 10;
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS AZURE POSTGRESQL QUERY RESULTS SPELL CHECKER

Results Messages Open in New Tab

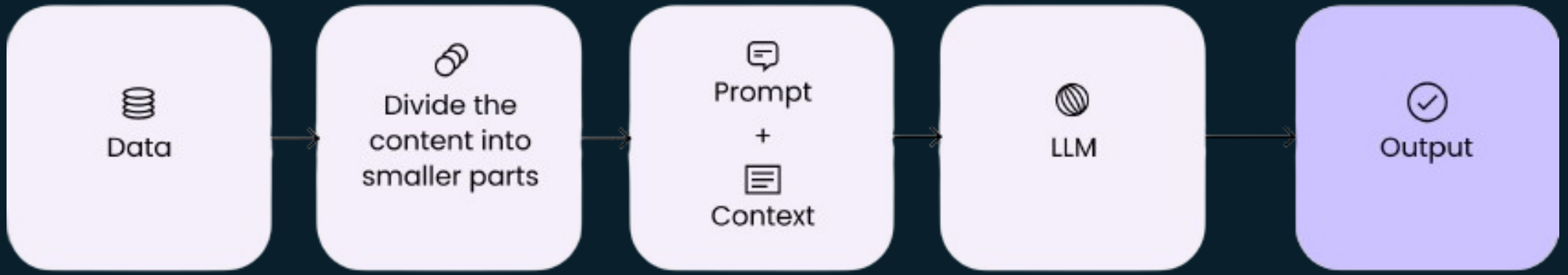
	id	name	opinion
1	615468	Le Vette v. Hardman Estate	"Morris, J.\nAppeal from an order of nonsuit and dism…
2	768356	Uhl Bros. v. Hull	"Holcomb, J.\nAppellant, subtenant of a former lessee…
3	674990	Woolworth Co. v. City of Seattle	"Mitchell, J.\nAppellant, plaintiff below, was engage…
4	4938756	Stevens v. King County	"Donworth, J.\nRespondents, as plaintiffs, brought th…
5	8848167	Wilkening v. Watkins Distributors, Inc.	"[As amended by order of the Court of Appeals October…
6	1346648	Tombari v. City of Spokane	"Millard, J.\nThis action was brought to recover for …
7	838633	Downie v. City of Renton	"Millaed, J.\n-This action was brought to enjoin the …
8	5041745	Frisken v. Art Strand Floor Coverings, Inc.	"Rosellini, J.\nThe respondent, Florence Frisken, is …
9	630224	Imperial Candy Co. v. City of Seattle	"Main, J.\nThe purpose of this action was to recover …
10	685636	City of Spokane v. Fisher	"Tolman, J.\nThe city of Spokane was made defendant i…

Agenda – Build AI Legal App on Postgres



RAG 101

RAG stands for **Retrieval-Augmented Generation**. It's a technique used to enhance the output of large language models (LLMs) by integrating external knowledge sources.



Two Problems in Information Retrieval

Problem #1: Scale

Efficiently scaling vector stores to 1M+ of vectors is hard.

Problem #2: Accuracy

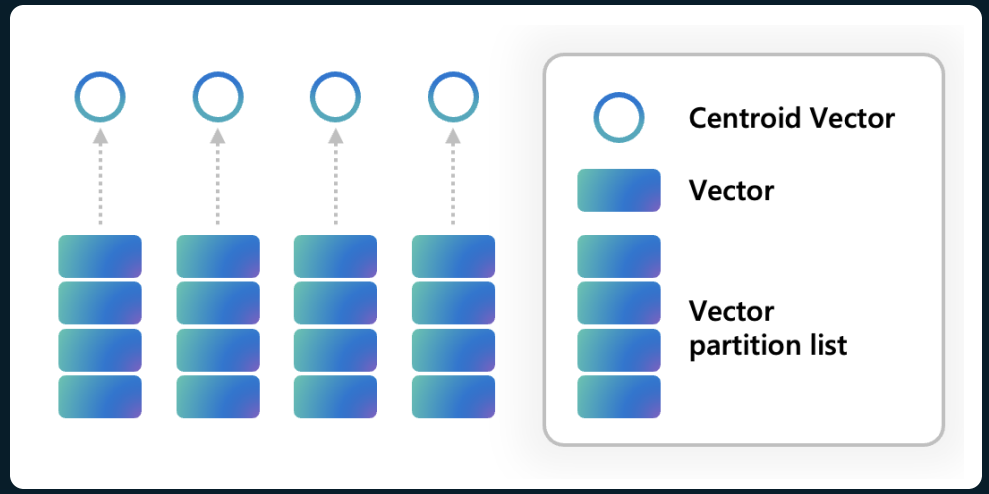
Quality of GenAI app responses and vector search accuracy need to improve.

Solving Problem #1 – Scale with Vector Indexing

Vector indexes popular today

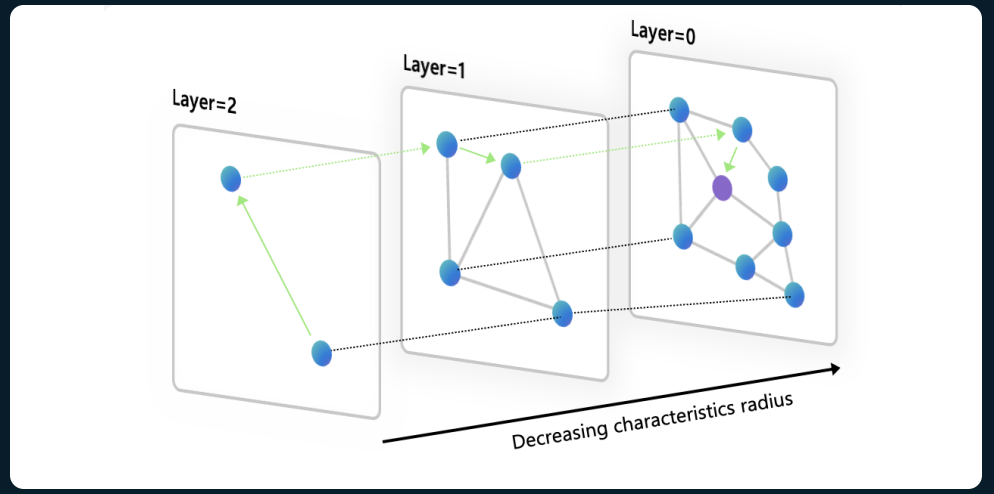
IVFFlat

- Clusters vectors by applying k-means clustering.
- Memory efficient but requires index rebuilds.



HNSW

- Builds a multi-layer graph with long and short connections between the vectors.
- The graph can be incrementally updated.



Vector indexes popular today

IVFFlat

- Clusters vectors by applying k-means clustering.
- Memory efficient but requires index rebuilds.

Generally speaking… IVFFLAT does its best work for larger datasets with low dimensionality.

HNSW

- Builds a multi-layer graph with long and short connections between the vectors.
- The graph can be incrementally updated.

Generally speaking… HNSW does its best work for smaller datasets with high dimensionality.

Vector indexes popular today

IVFFlat

- Clusters vectors by applying k-means clustering.
- Memory efficient but requires index rebuilds.

Generally speaking… IVFFLAT does its best work for larger datasets with low dimensionality.

HNSW

- Builds a multi-layer graph with long and short connections between the vectors.
- The graph can be incrementally updated.

Generally speaking… HNSW does its best work for smaller datasets with high dimensionality.

What do you do for large datasets with high dimensionality?

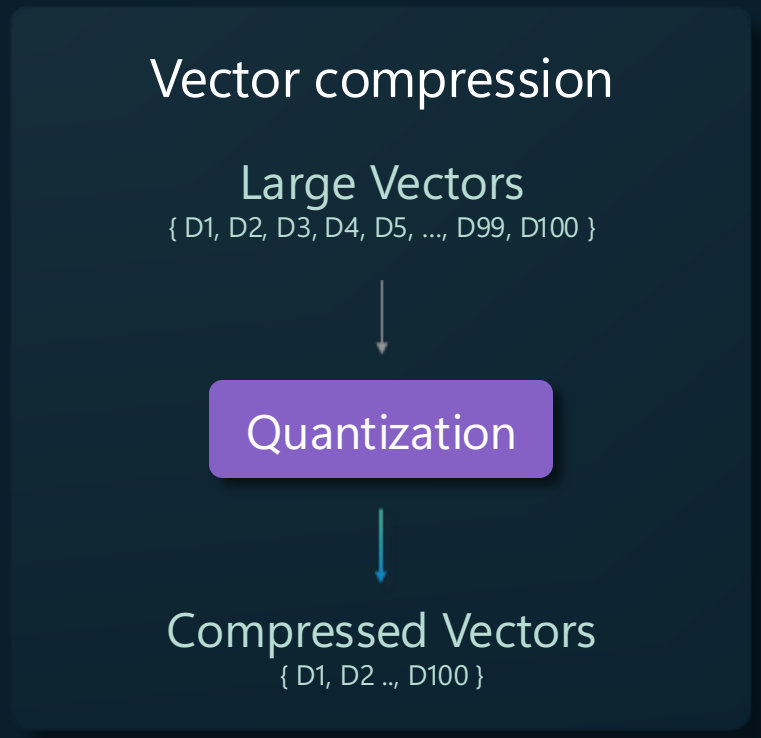
DiskANN Vector Index

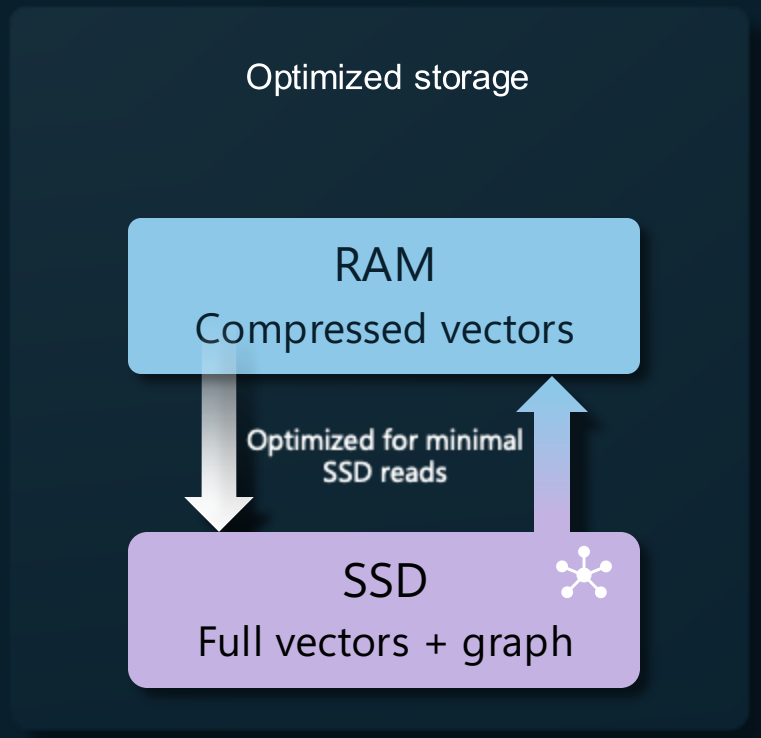
Highly performant, scalable, and accurate index for vectors

Superior to IVFLAT and HNSW

Reduced memory footprint with quantization and storing SSD

Huge **cost savings** at scale due to reduced memory footprint.





DiskANN vs HNSW index size

```
postgres=# \dt+ cases_*
                                    List of relations
 Schema |       Name       | Type  |  Owner   | Persistence | Access method |  Size   | Description
--------+------------------+-------+----------+-------------+---------------+---------+-------------
 public | cases_diskann    | table | postgres | permanent   | heap          | 6276 MB |
 public | cases_hnsw       | table | postgres | permanent   | heap          | 6276 MB |
 public | cases_metadata   | table | postgres | permanent   | heap          | 13 GB   |
 public | cases_no_index   | table | postgres | permanent   | heap          | 6276 MB |
 public | cases_playground | table | postgres | permanent   | heap          | 3896 kB |
(5 rows)

postgres=# \di+ cases_*
                                                List of relations
 Schema |                 Name                 | Type  |  Owner   |     Table      | Persistence | Access method |  Size   |
--------+--------------------------------------+-------+----------+----------------+-------------+---------------+---------+-
 public | cases_diskann_description_vector_idx1 | index | postgres | cases_diskann  | permanent   | diskann       | 150 MB  |
 public | cases_hnsw_description_vector_idx     | index | postgres | cases_hnsw     | permanent   | hnsw          | 3688 MB |
 public | cases_metadata_pkey                   | index | postgres | cases_metadata | permanent   | btree         | 207 MB  |
 public | cases_pkey                            | index | postgres | cases          | permanent   | btree         | 2 MB    |
(4 rows)
```

Using DiskANN

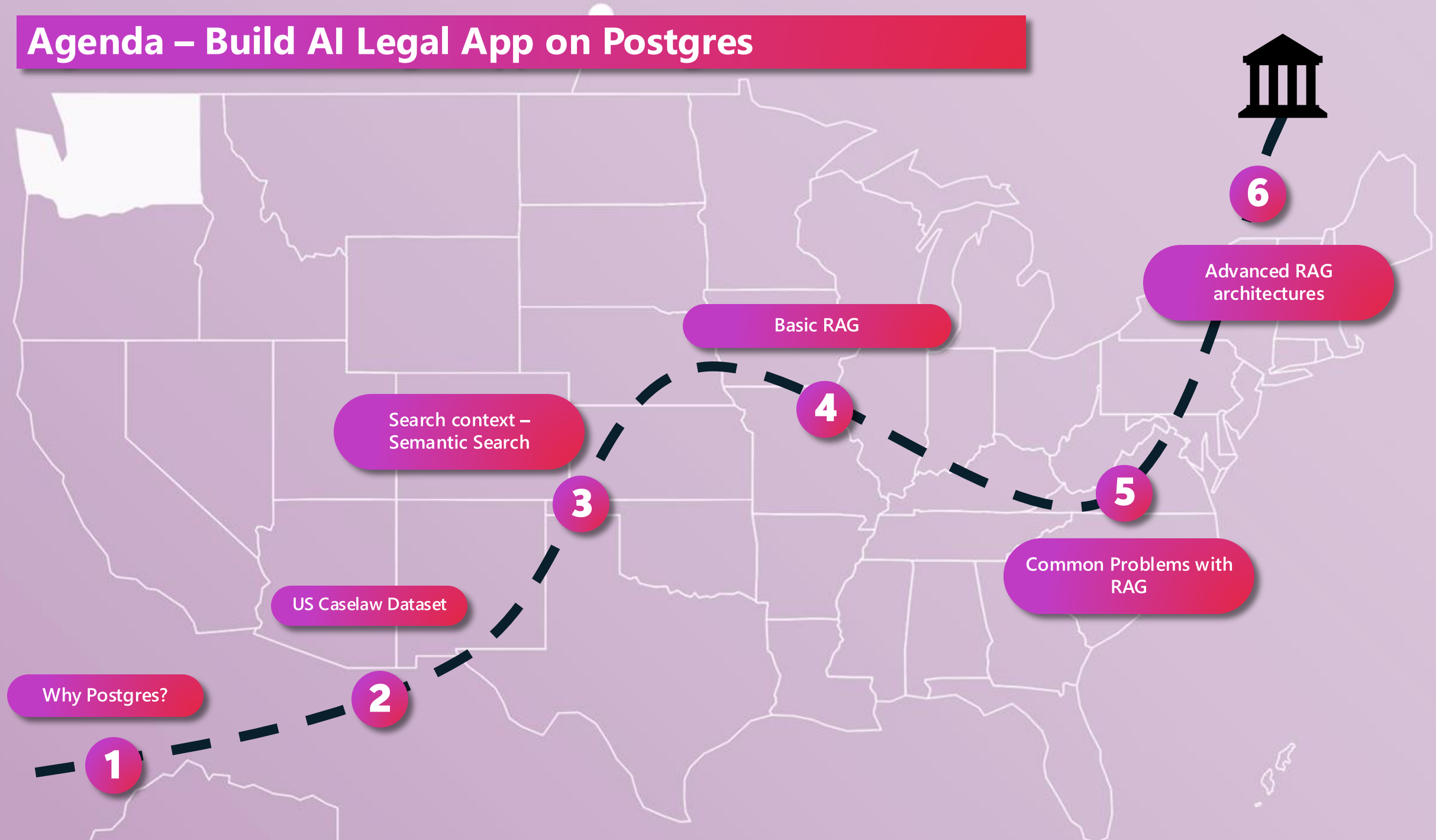
- Now, let's dive into DiskANN for your vector data
- We can create an index on the vector column, so we can do faster searching on the vector data.

```
CREATE INDEX ON cases USING diskann
(opinion_vec vector_cosine_ops)

-- Create DiskANN index to improve search
times
```

Solving Problem #2 – Accuracy with Advanced RAG architectures

Agenda – Build AI Legal App on Postgres



Improving Accuracy of GenAI apps

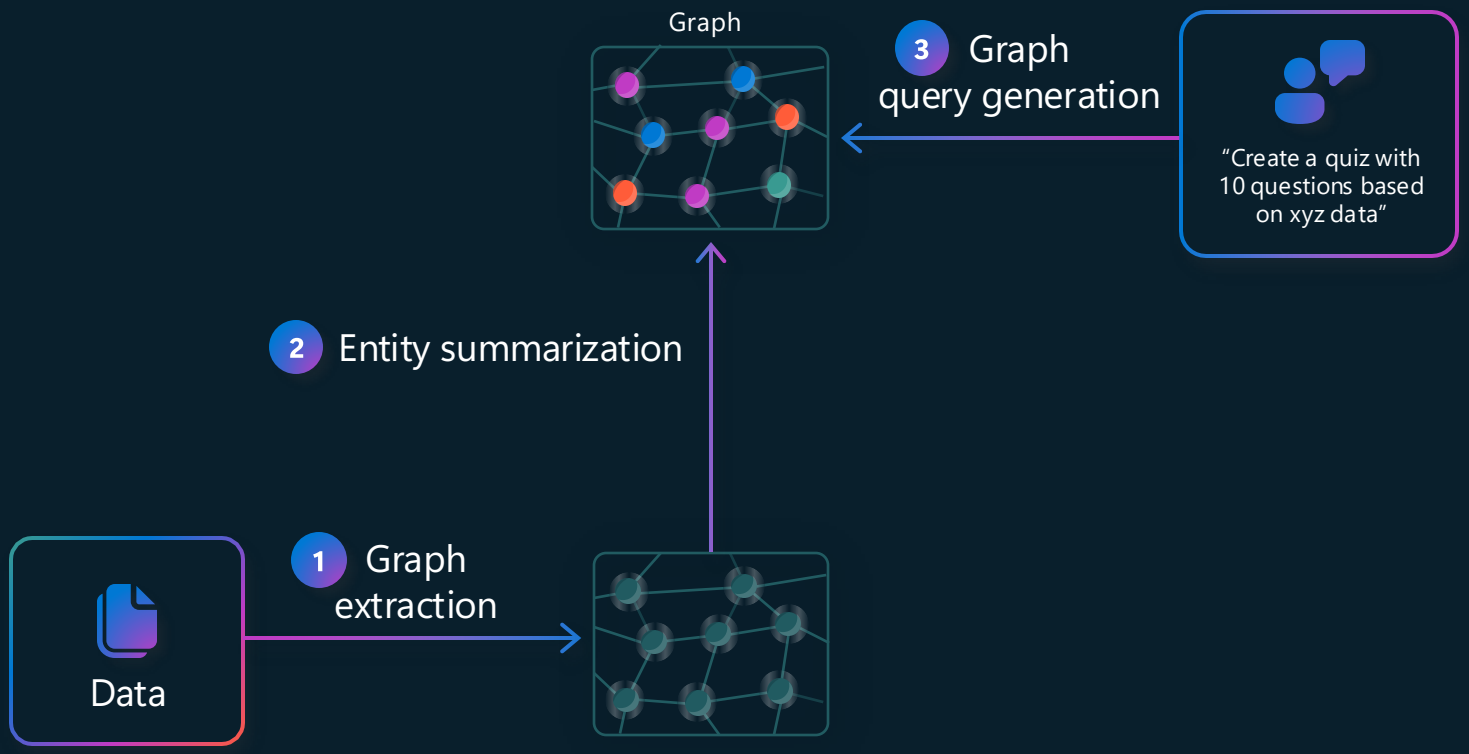
Basic RAG

- Chunking strategies
- Bigger embeddings
- Query rewriting
- Hybrid search
- Metadata filtering

Advanced RAG

- Semantic Ranking
- Hierarchical summarization (RAPTOR)
- Knowledge graphs (GraphRAG)
- Agentic systems

GraphRAG
Overview



Project GraphRAG: aka.ms/graphrag

GraphRAG Steps

		Legal Research Copilot Solution
1	Graph extraction	→ Pre-extracted citation graph
2	Entity summarization	→ Entity summarization
3	Graph query generation at query time	→ Specialized graph query

Apache AGE

Graph database extension for PostgreSQL

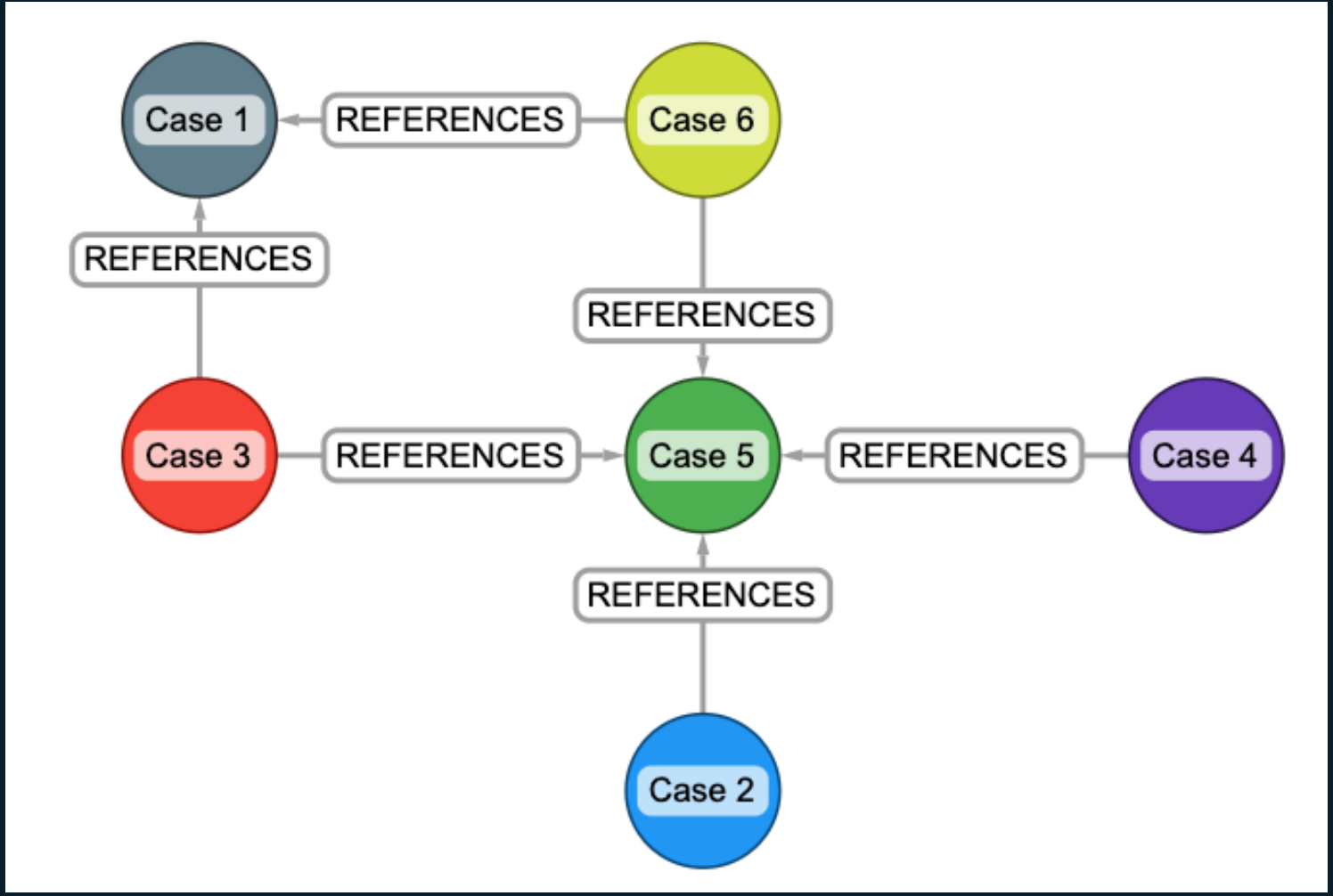
Graph database
Plugin for PG

Cypher + SQL
Hybrid Queries

Fast Graph
Processing

Graph Visualization
& Analytics

Example Constructed Case Graph



Advanced Retrieval Technique - GraphRAG

- GraphRAG uses knowledge graphs to **improve results** in **complex scenarios,** by leveraging relationships
- For example, “Water leaking into the apartment from the floor above. What are the **prominent legal precedents** in Washington on this problem?”

```
graph AS (
  SELECT * from vector_search_ranked
  JOIN cypher('case_graph', $$
    MATCH ()-[r]->(n)
    RETURN n.case_id, COUNT(r) AS refs
  $$) as graph_query(case_id TEXT, refs
BIGINT)
  ON vector_search_ranked.id =
graph_query.case_id
)
```

```
62  graph AS (
63      SELECT graph_query.refs, vector_search_ranked.vector_rank, vector_search_ranked.*, grap
64      LEFT JOIN cypher('case_graph', $$
65              MATCH ()-[r]->(n)
66              RETURN n.case_id, COUNT(r) AS refs
67          $$) as graph_query(case_id TEXT, refs BIGINT)
68      ON vector_search_ranked.id = graph_query.case_id::int
69  ),
70  graph_ranked AS (
71      SELECT RANK() OVER (ORDER BY COALESCE(graph.refs, 0) DESC) AS graph_rank, graph.*
72      FROM graph ORDER BY graph_rank DESC
73  ),
74  rrf AS (
75      SELECT
76          COALESCE(1.0 / (60 + graph_ranked.graph_rank), 0.0) +
77          COALESCE(1.0 / (60 + graph_ranked.vector_search_rank), 0.0) AS score,
78          graph_ranked.*
79      FROM graph_ranked
```

Use RRF to combine Graph and Reranked results.

PROBLEMS 2 | OUTPUT | DEBUG CONSOLE | TERMINAL | PORTS | POSTGRESQL QUERY RESULTS

Results | Messages | Open in New Tab

	id	case_name	vector_search…	graph_rank
1	1279441	Tope v. King County	6	13
2	4975399	Laurelon Terrace, Inc. v. City of Seattle	12	7
3	615468	Le Vette v. Hardman Estate	1	23
4	1034620	Jorgensen v. Massart	2	23

Legal Research Copilot

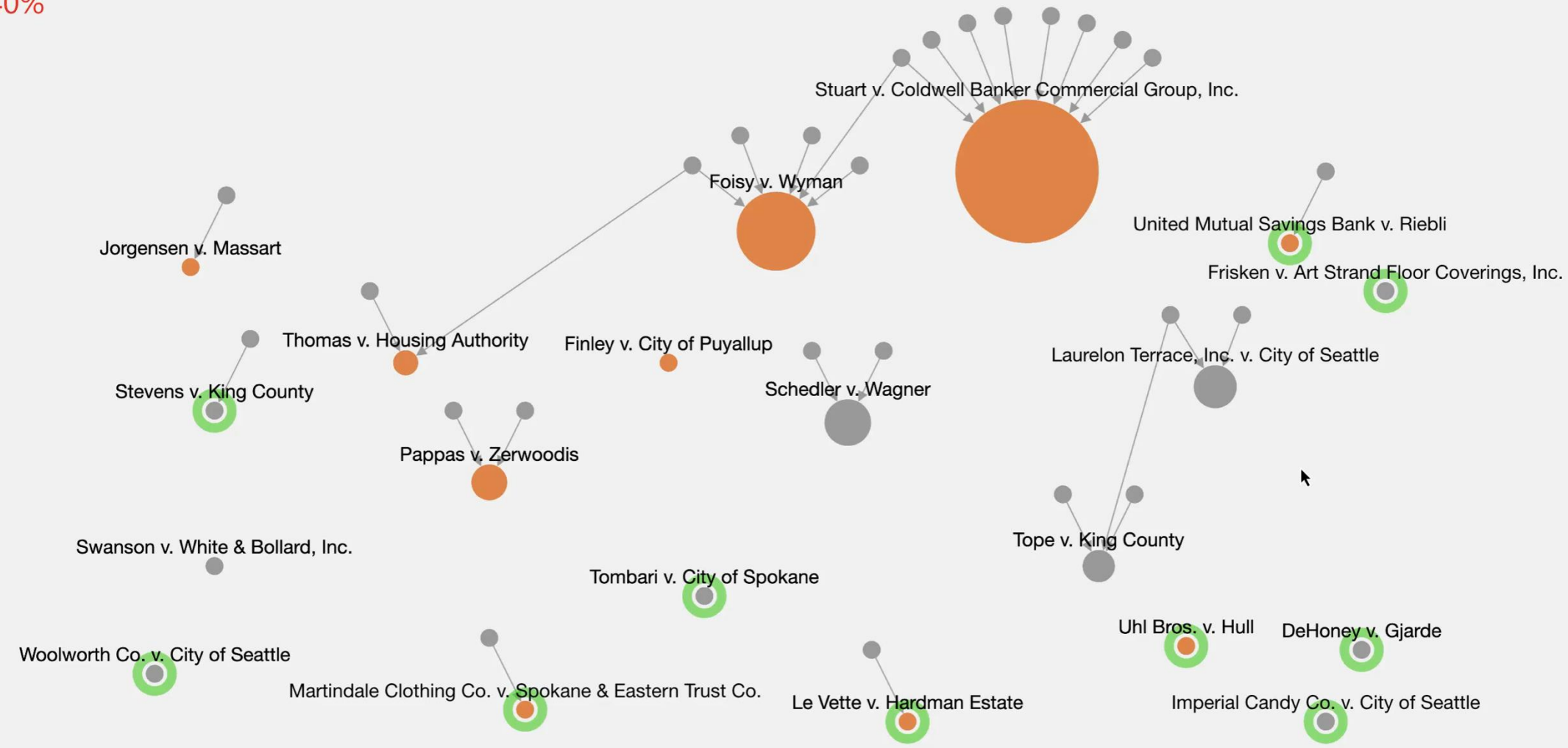
Vector Search | Semantic Ranker | GraphRAG + Semantic Ranker

Demo: GraphRAG on PostgreSQL

Clear

Citation Graph

Recall: 40%



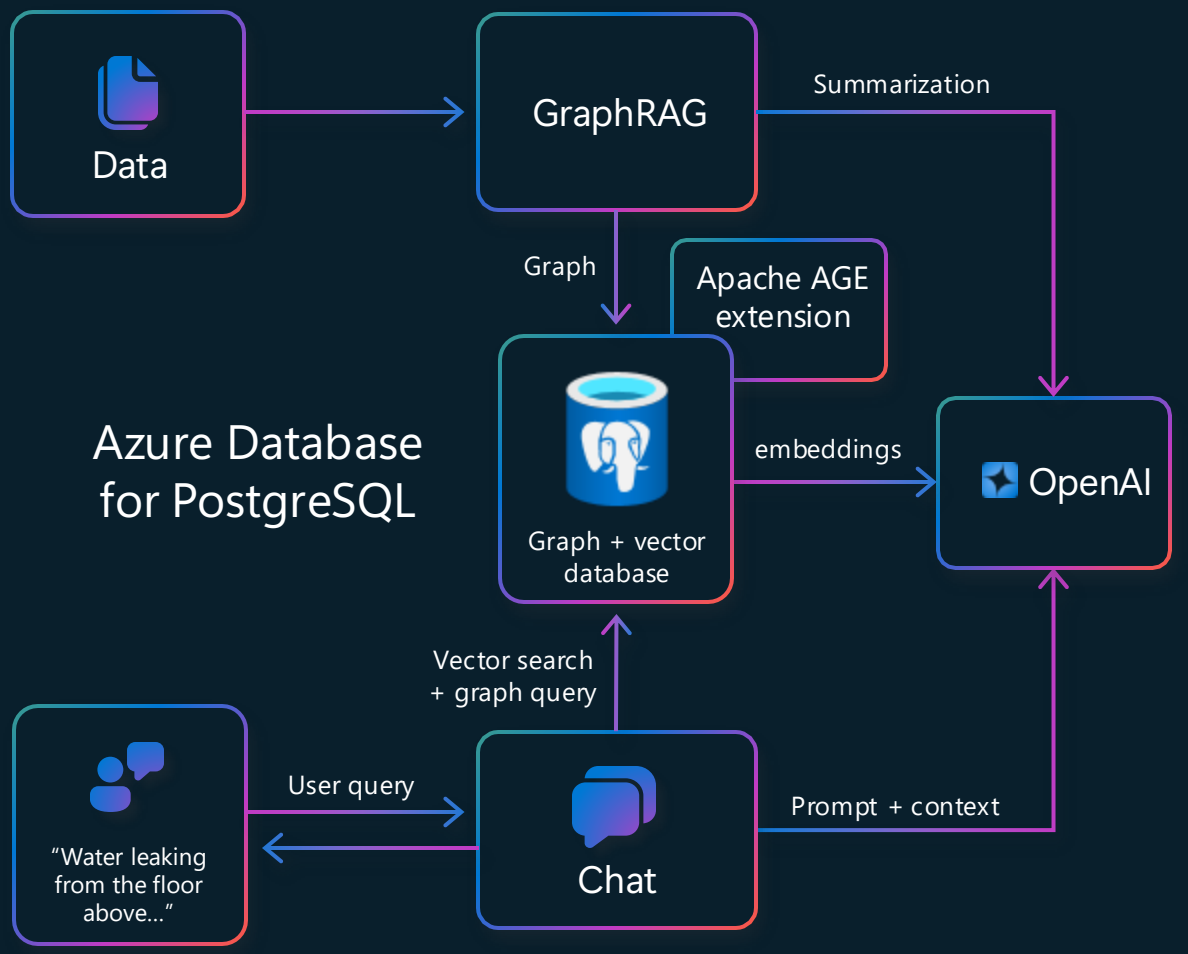
GraphRAG Solution Accelerator for Postgres

Overview

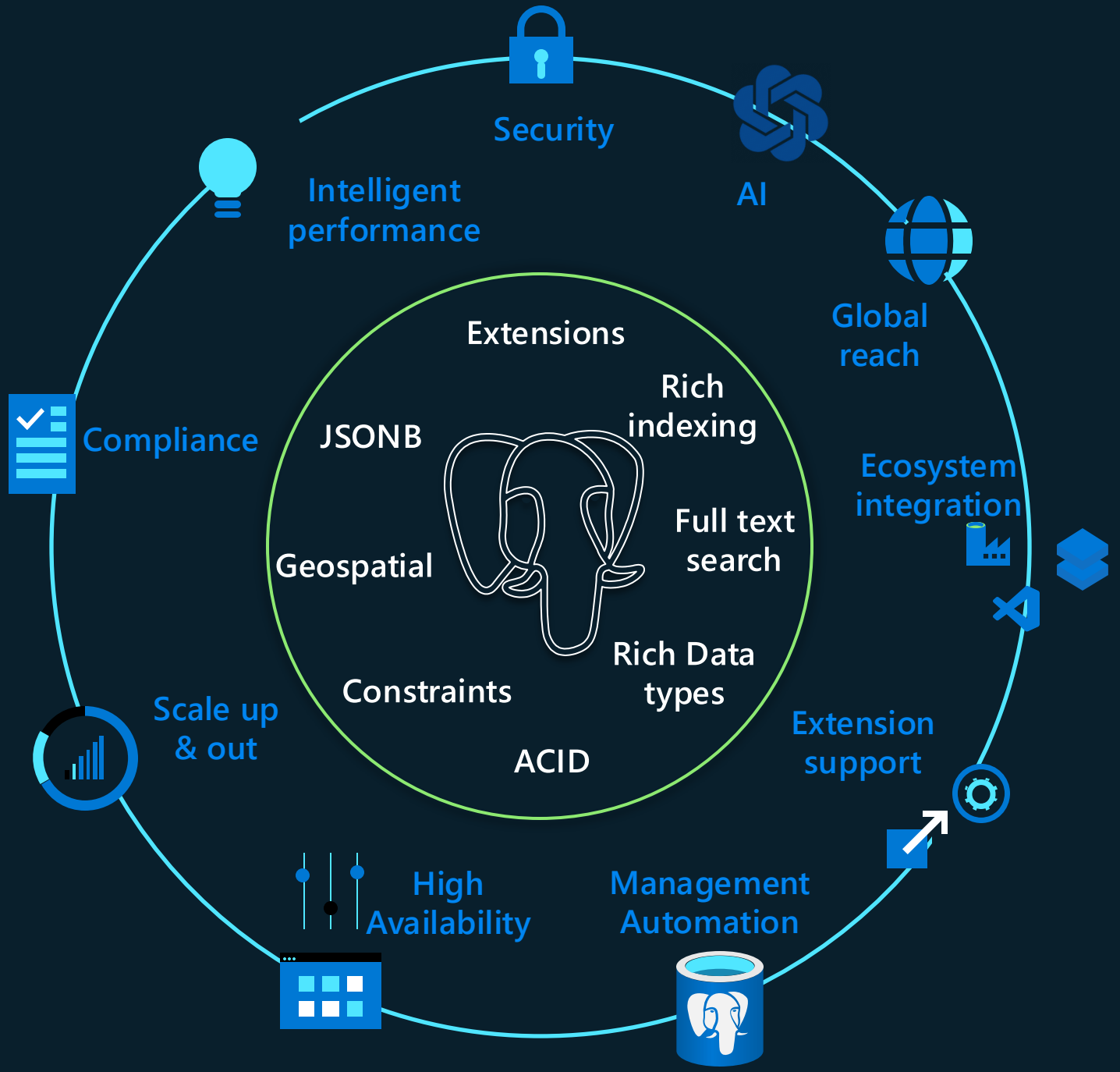
- Legal Research Copilot app
- U.S. Case Law dataset (0.5 million cases)

Available Now!

- Blog: aka.ms/pg-graphrag
- Repo: aka.ms/pg-graphrag-repo



Azure Database for PostgreSQL: AI-Ready for Enterprise Applications



More Resources

Azure Postgres and AI Agents

- aka.ms/pg-ai-agents-blog

Postgres AI Solution Accelerators

- aka.ms/postgres-semantic-ranker
- aka.ms/postgres-graphrag-solution
- aka.ms/pg-byoac-docs

Read the Azure Postgres blog

- aka.ms/azurepostgresblog