

# Can PostgreSQL have a more prominent role in the AI boom?

*"Maybe in 20 years,  
everything will be done  
in PostgreSQL?"*

*(Simon Riggs,  
talk on PG conf EU 2023)*

- Josef Machytka - NetApp
- 2024-06-27 Swiss PG day



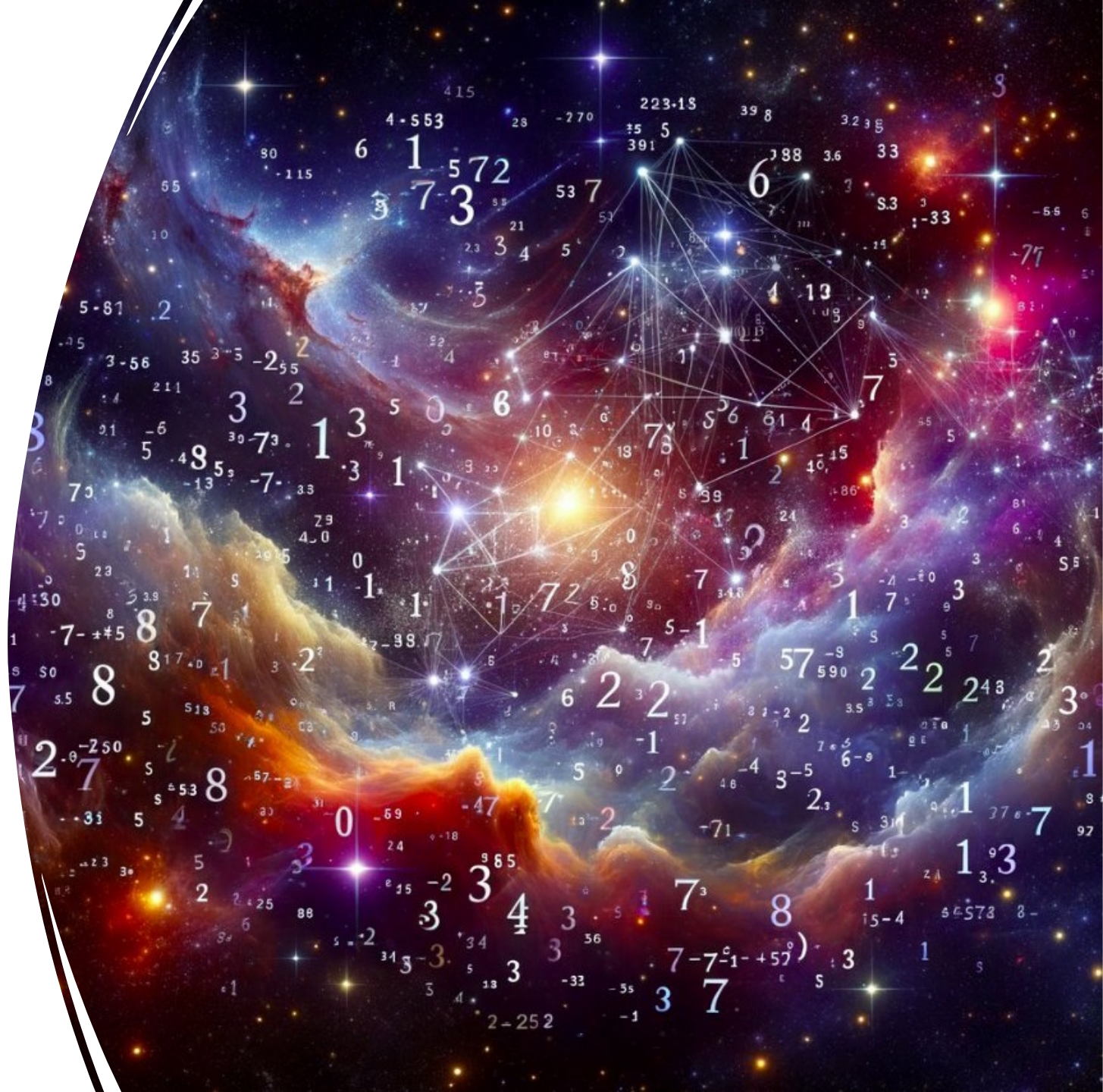
*Pictures generated  
with DeepDreamGenerator*



# PostgreSQL already has vector storage and vector search

---

- Extension pgvector, with vector data type, indexing methods and vector search.
- We have seen very nice talks about it here. I really enjoyed them!
- But every database is now advertising vector search...
- Can we do even better in PostgreSQL?





# All the magic is happening inside an embedding model

---

- Embedding models are algorithms trained to encapsulate information into dense representations in a multi-dimensional space.

([aws.amazon.com/what-is/embeddings-in-machine-learning](https://aws.amazon.com/what-is/embeddings-in-machine-learning))

- Vector embeddings are a way to convert words and sentences and other data into numbers that capture their meaning and relationships.

([www.elastic.co/what-is/vector-embedding](https://www.elastic.co/what-is/vector-embedding))

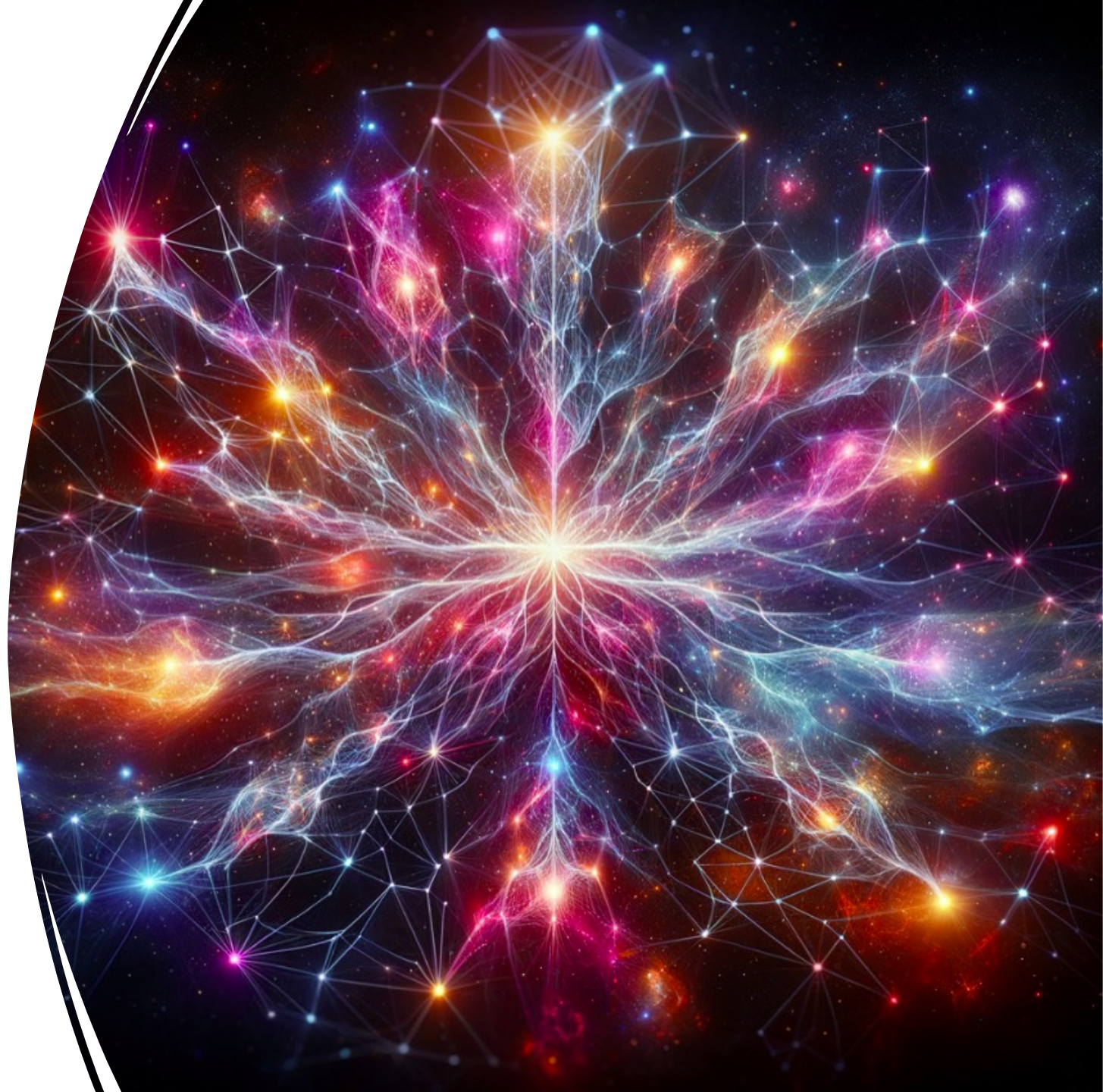




# Neural networks and ML models are all about vectors and numbers

---

- Whole AI model is just a very advanced statistics.
- Stores data in its own binary format.
- TensorFlow ML model:
  - meta data of the graph structure
  - variables in checkpoint files (saved steps during training of the model)
  - index file
- They internally simulate graph database; they could easily hit different limits and performance problems.
- From big data perspective it is just another bunch of vectors. So maybe with PostgreSQL they can do better?





# Databases obtain new functionality for ML, so why not AI?

---

- BigQuery already implemented ML and AI related functionality - you can use pretrained models directly in SQL
- PostgresML implements ML functionality
- New extensions for ML and AI models?
- Much more dimensions would be necessary – AI uses internally even billions of dimensions...
- I am not the ML expert; I am just asking...
- But maybe future really is more about PostgreSQL being everywhere...
- Could be also a very good new impulse for PostgreSQL improvements...

