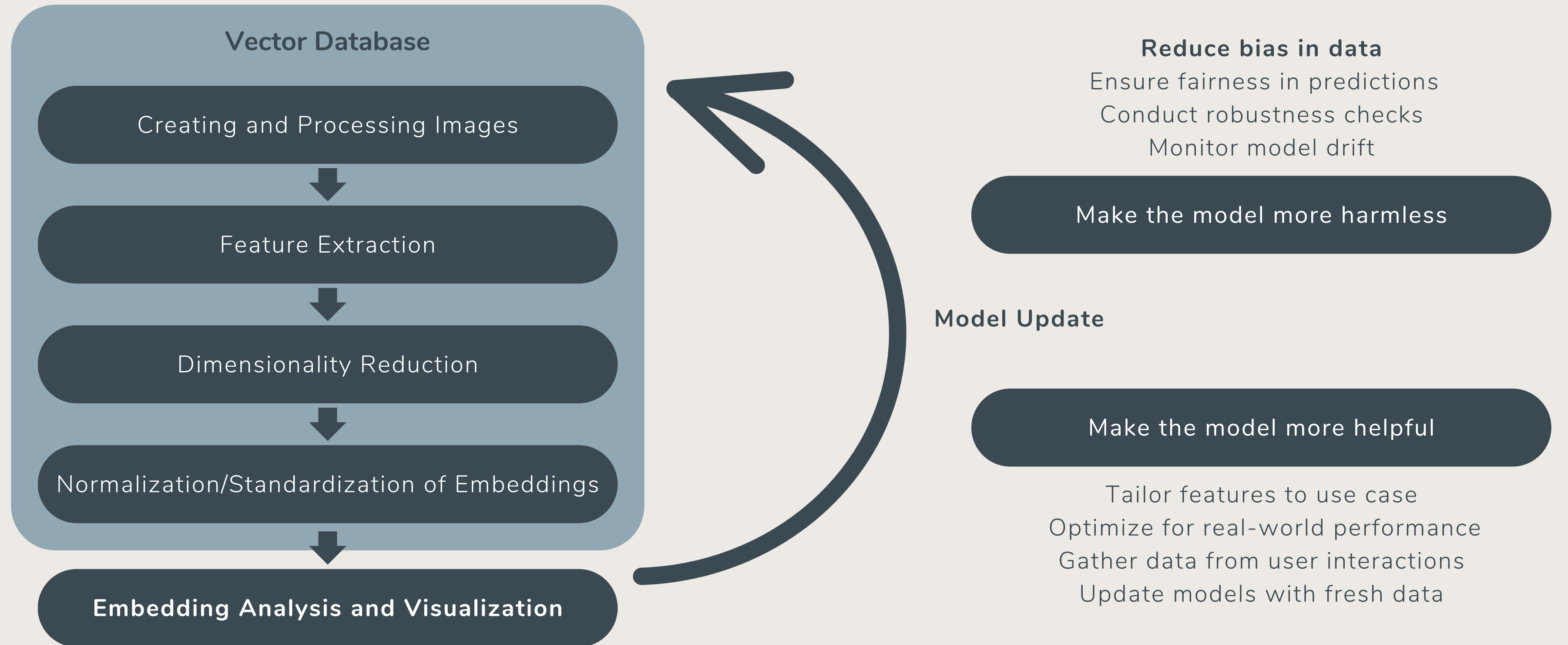


Supporting Human-in-the-Loop Evaluation
of Biased Image Embeddings
in Vector Databases
through Configurable Visualizations

AUTHOR:

Ankita Suresh Shanbhag
University of California, Berkeley

THE PROBLEM



Example - a marketing firm stores LLM-generated human face images for various occupations in a vector db. The data engineer wants to identify and address the 'biggest' biases - eg. images of nurses are disproportionately female-presenting.

CHALLENGES IN CURRENT APPROACHES

Complexity of Bias



Current algorithms often fail to grasp the nuanced cultural, societal, and historical contexts of biases

Limitations in Automation



Difficulty in automating bias visualization due to the subtle nature of most biases

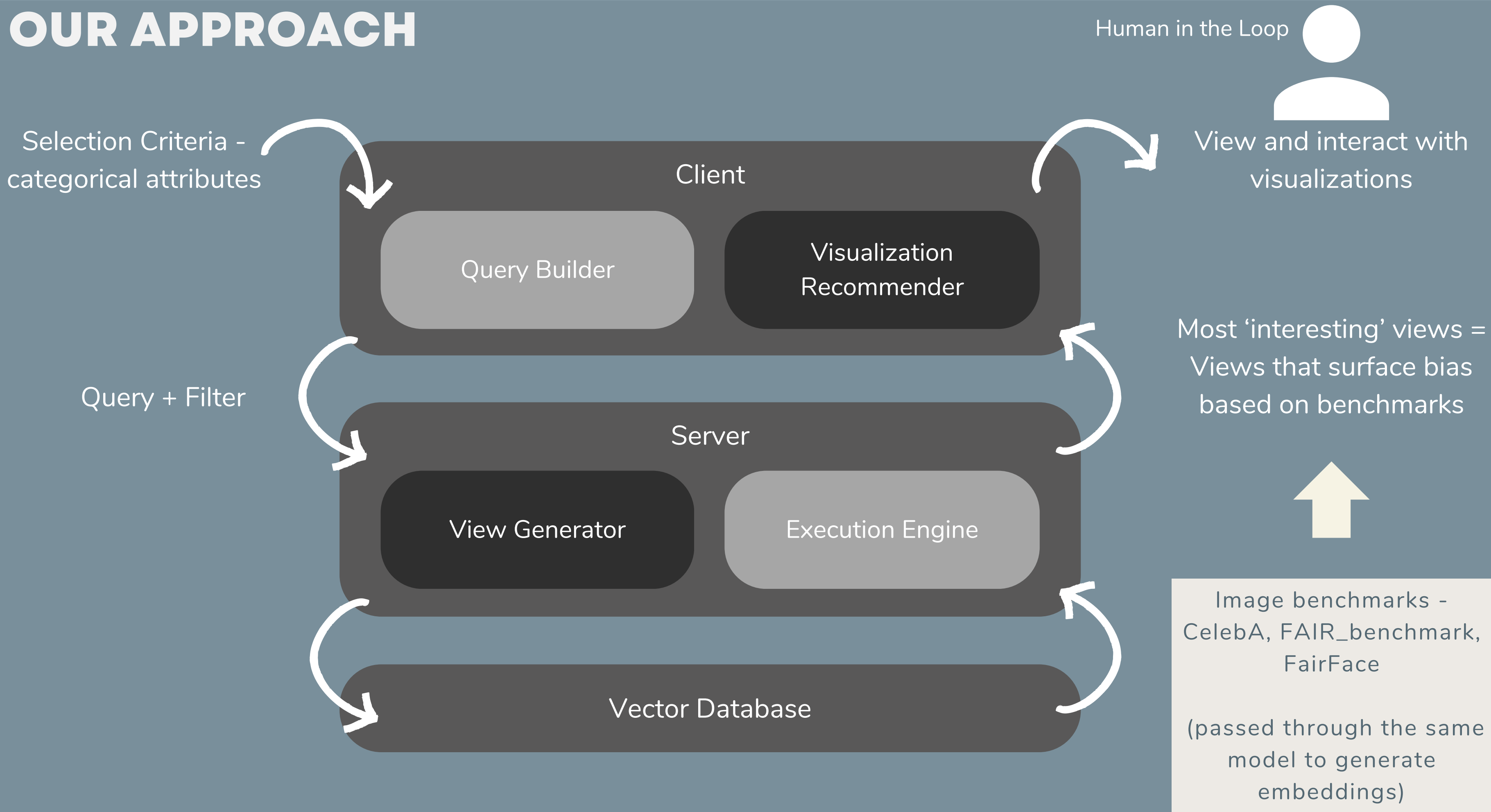
Inadequacy of Existing Tools



Lack of comprehensive tools for effective visualization and interpretation of biases in image embeddings

Vector databases are only recently gaining traction, but at a rapid pace. Precedents exist for traditional databases - eg. SeeDB.

OUR APPROACH



BIAS AS THE HEURISTIC



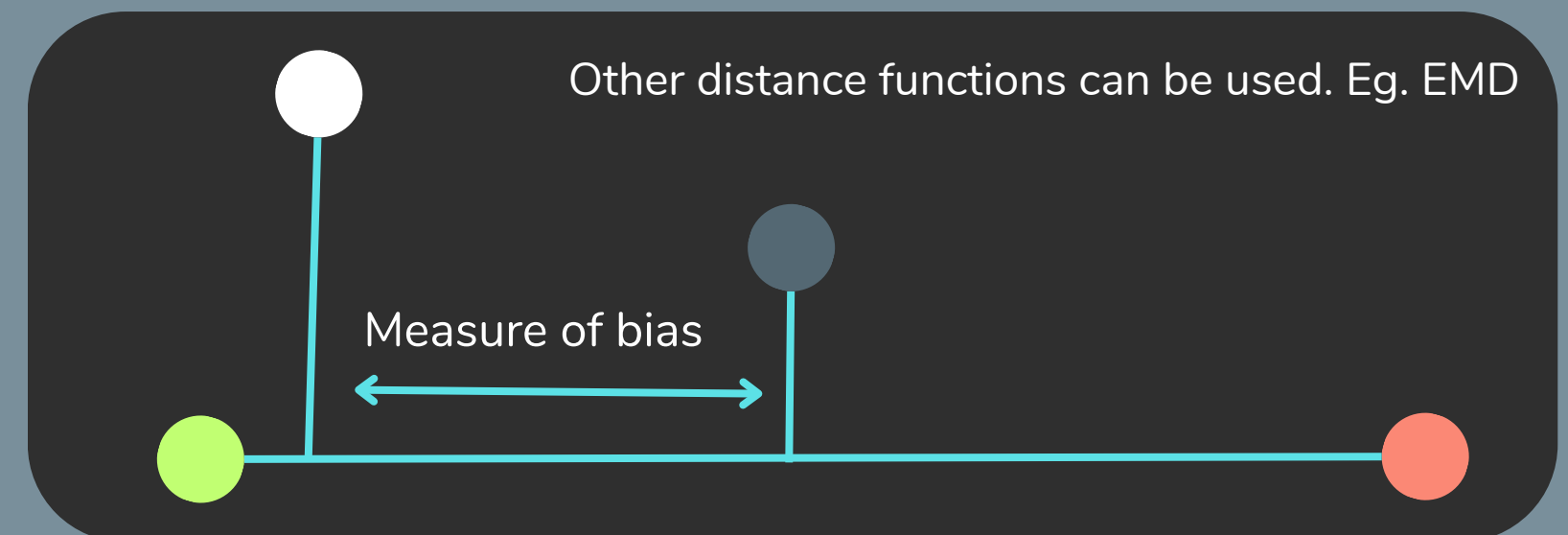
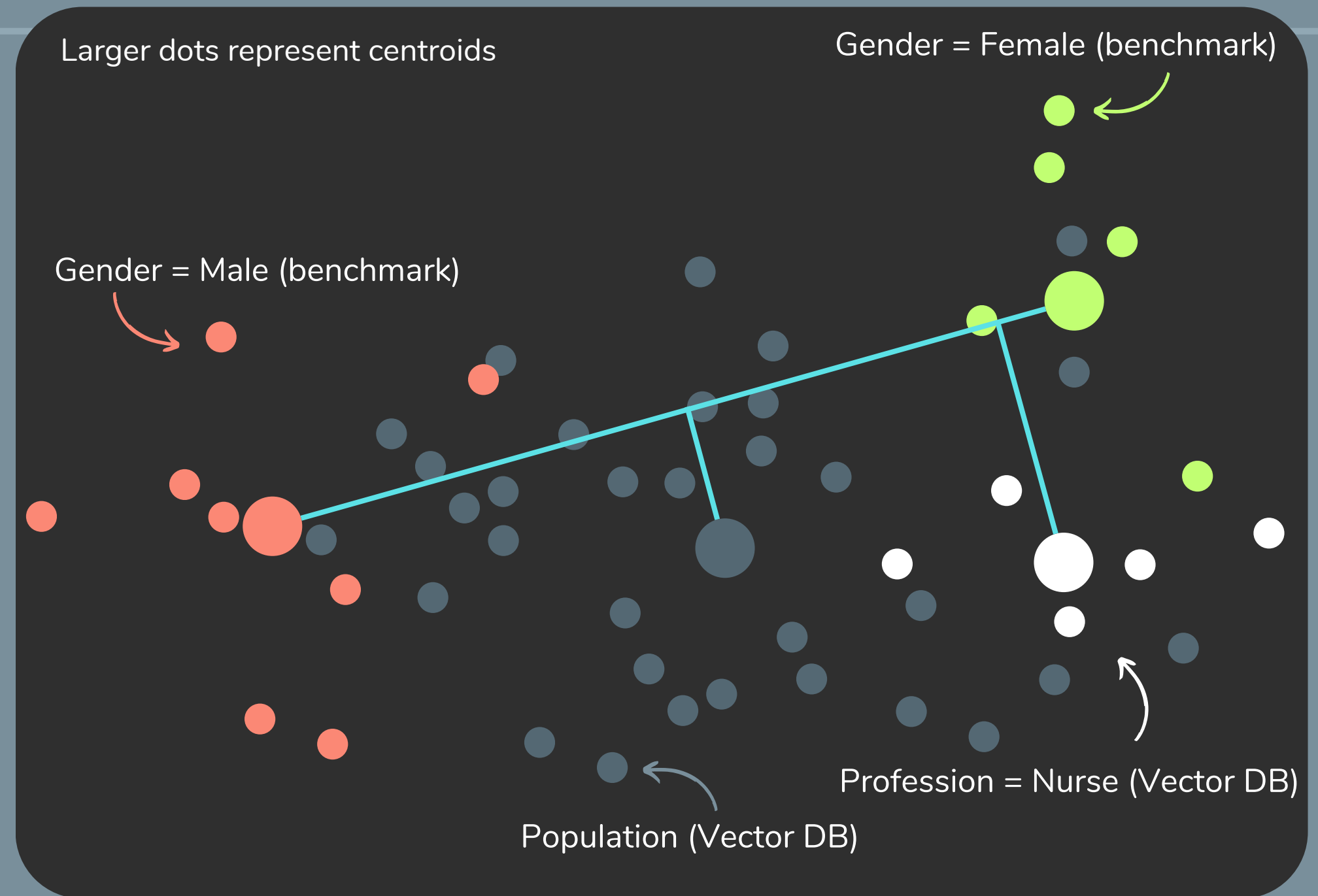
Example image in the vector db

Values - vector embedding of n dimensions
Metadata - image attribute (eg. profession)



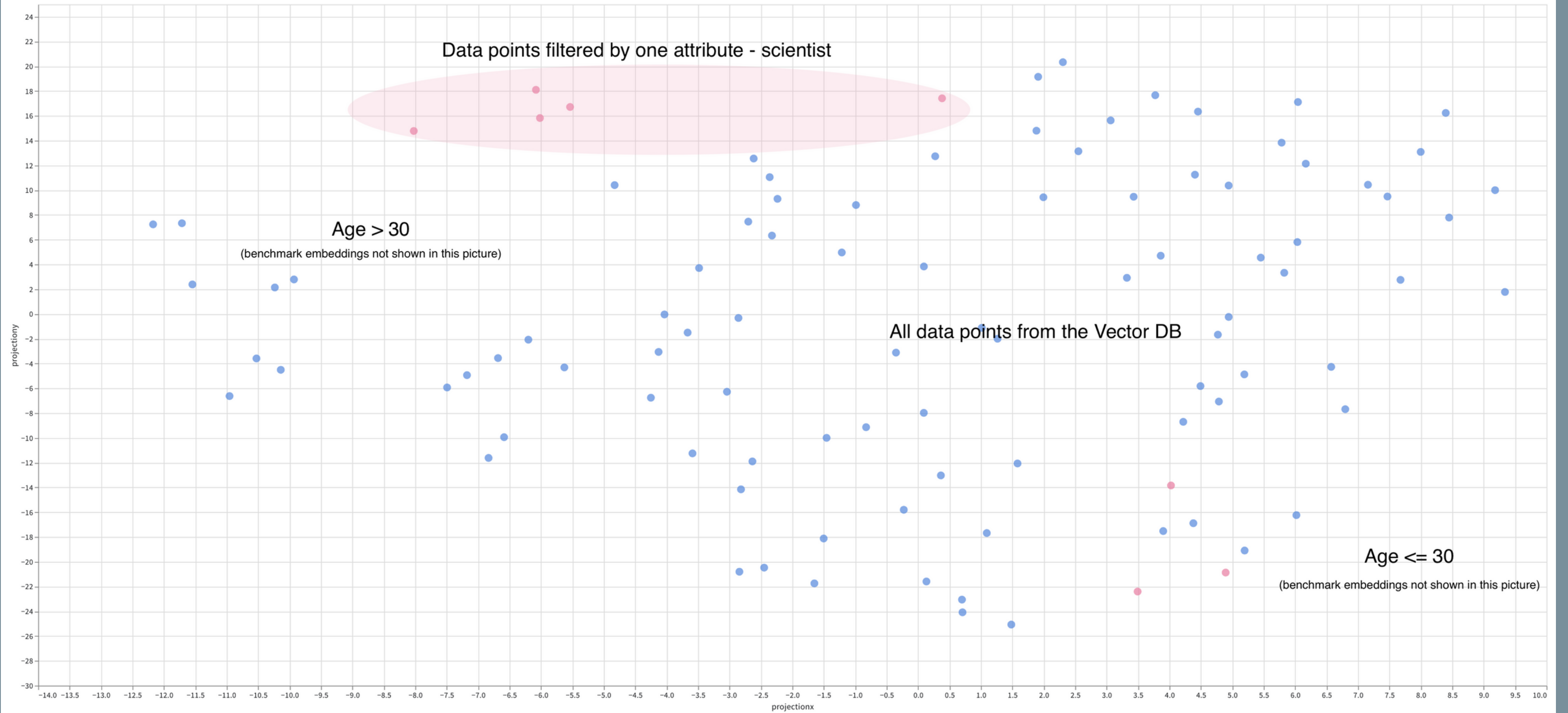
Example benchmark

Values - vector embedding of n dimensions
Metadata - attribute (eg. gender), class (eg. female)



EXAMPLE FROM WANDB (USING PCA)

`runs.summary["embeddings_table"]`



EVALUATION

Baseline

Vector database with Weights & Measures connection set up and benchmark embeddings provided (in the same or different db)

After

Baseline + visualizations sorted by bias + user control over selection of attribute filters

Questions asked of the human tester

- Which of the two did you find more helpful?
- Which of the two did you find more intuitive?
- Which of the two helped you derive insights from this data (if at all) faster?

EXTENSIONS AND FUTURE WORK

- **Performance Optimization:** Improving the system's performance - using an execution engine
- **Enhanced Visualization Tools:** Refining visualization tools for a more intuitive and immersive user experience
- **Comparative Bias Analysis:** Introducing a framework for comparative studies across multiple bias dimensions - using a query builder
- **Tuning Parameters:** Varying some of the parameters, such as the distance function, and evaluating their impact
- **Expanding the Problem Space:** Analyzing vector databases using other 'interestingness' functions (such as fairness and toxicity) and benchmarks
- **Evaluating Robustness against Benchmark Embedding Models:** Validating the effect (if any) of using a different model to generate benchmark embeddings

*Thank
you!*