

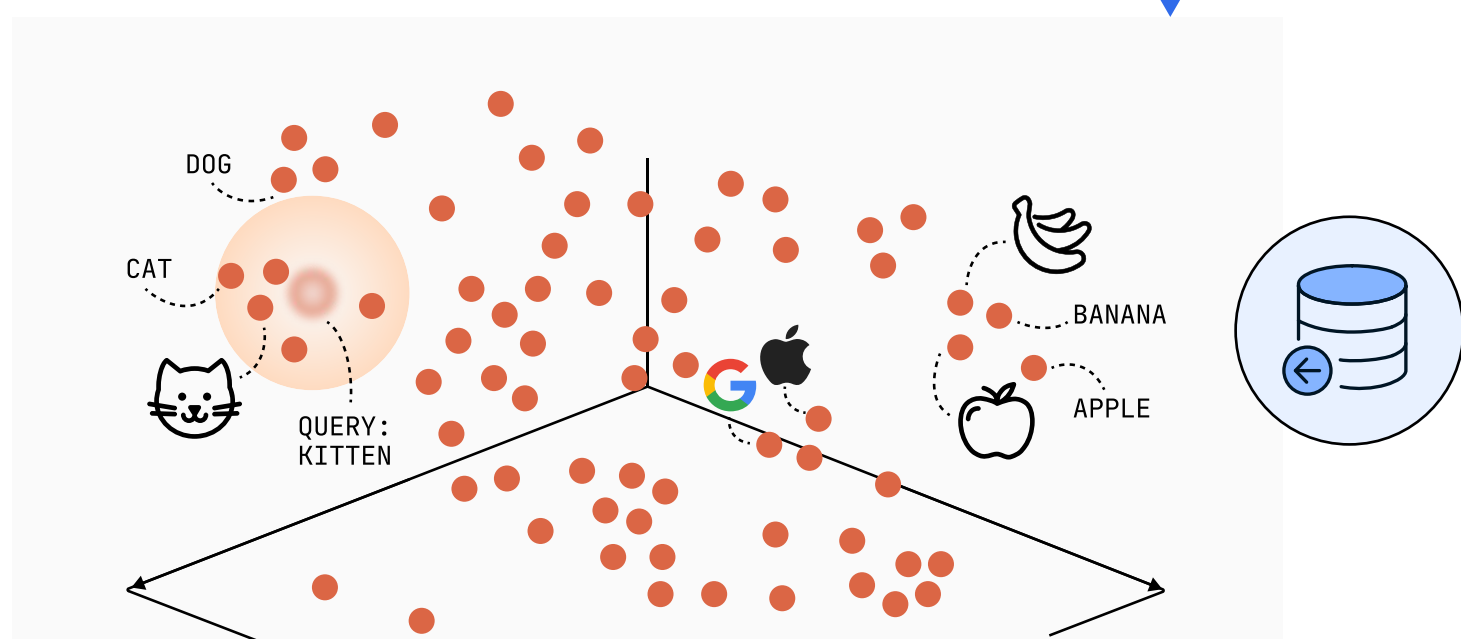
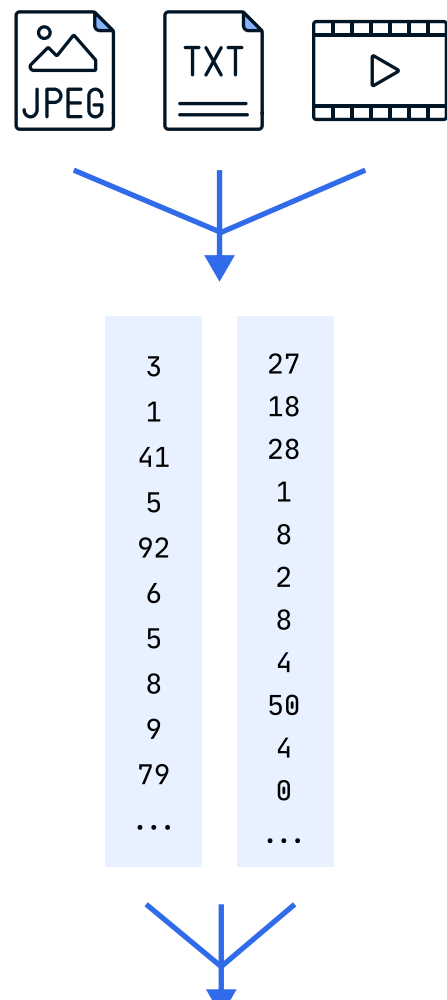
Using generative AI with different kinds of data



Generative AI is most commonly associated with unstructured data such as raw text, images, video, audio and all other kinds of media.

To train or augment a generative AI model, data from these media is converted into multidimensional numerical representations called vectors, which can be used to determine similarity between entities of all kinds. Vectors are stored in special databases called vector databases.

The illustration shows how data is turned into long strings of numbers that correspond to positions in an n-dimensional space (represented by three dimensions for the sake of tractability).



Why use semi-structured data?

Much of the data containing details about an organization's operations is structured and stored in tables, not as media files. However, these tables may still contain rich textual information that can be used to power chatbots and other products that depend on knowledge about your organization's operations. This data is semi-structured, being stored in a tabular format but containing free-form text.

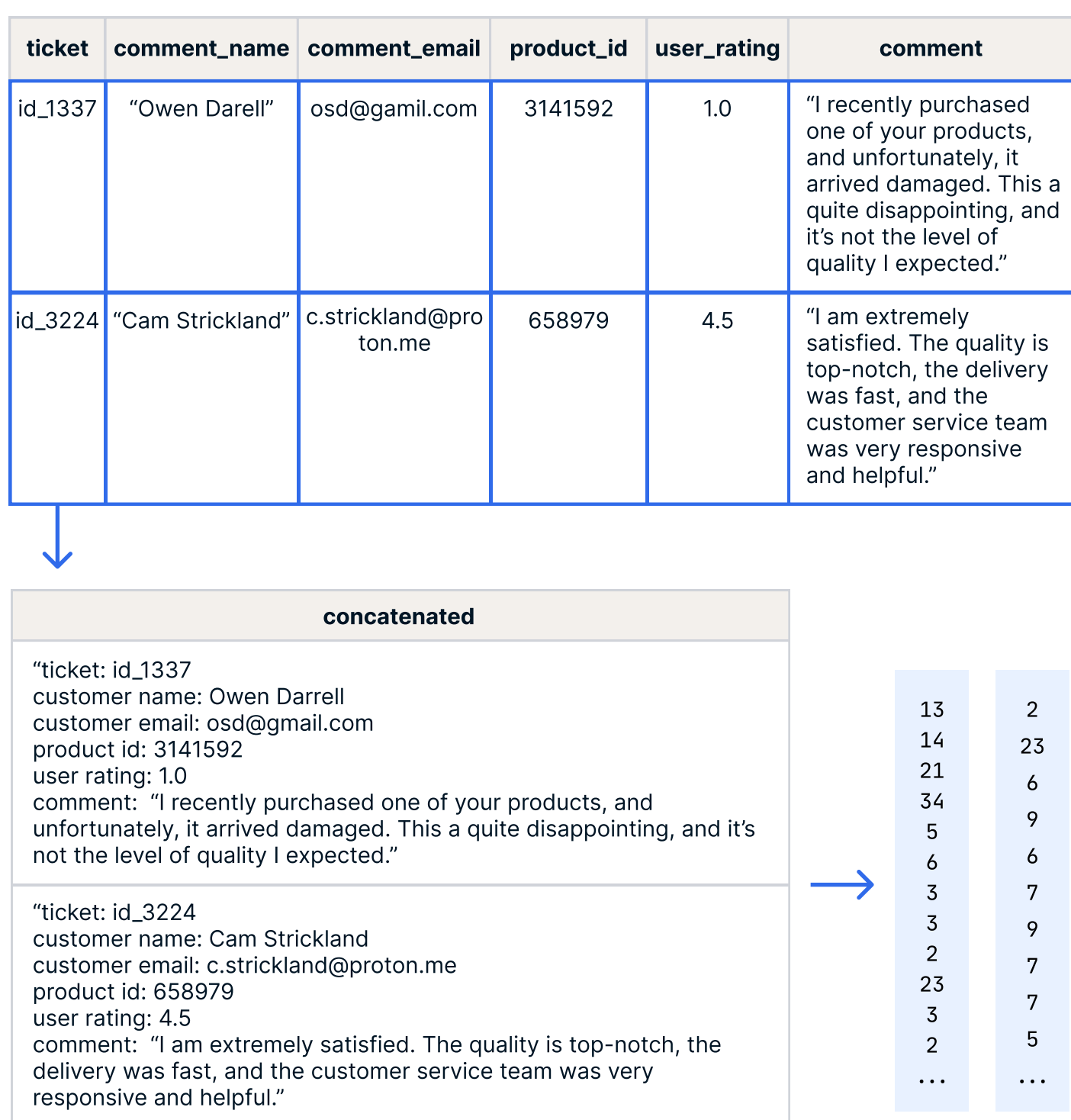
Extract text from text-rich tables

You can directly extract and vectorize the contents of text-rich fields from tables to provide extra context.



Extract, concatenate and add context

Another method is to concatenate different fields together to build a text-rich field, and then extract and vectorize it. You may add additional context as well.



What about structured data?

Critical metrics and facts about your organization, specific accounts for customers and prospects, team performance, business outcomes and more are usually stored in a structured, relational format. Traversing all of this data using a business intelligence platform or SQL can be a painful exercise for analysts and stakeholders. There is a strong case for the ability to interact with this data using natural language.

One way to use natural language to interact with important metrics and facts about your organization is through text-to-SQL (or text-to-Python, or other languages) business intelligence tools.

