Some questions before we proceed:

1. Is the text coming from Sefaria?
2. Am I right that the database we have now is just for testing purposes, and that the future version will include more fields?
3. Does the current database structure allow more than one figurative language example to be detected per verse?
4. How might we design the database to ensure that in the future we can identify a given metaphor and then identify a) how many other metaphors are in the same verse; b) all metaphors within 1 verse before or after that verse; c) all metaphors within 5 verses before or after that verse, etc (so that we can create a network map relating metaphors to each other by distance)?
5. I’d also like to make sure that we have a field that clarifies WHAT the metaphor is and WHY the LLM believed that was a metaphor (or simile, etc.)
6. In the current pipeline, what LLM is doing the labeling? How can we ensure that we stay within our usage limit and/or that if we hit our limit, we can switch to another LLM (e.g. Gemini)?
7. I’d like to try out our pipeline on Deuteronomy chapter 30 – it has lots of figurative speech.
8. I see we capture personification, simile and metaphor. Let’s also try to capture idiom and hyperbole.
9. It would be great to further subclassify the simile and metaphor into categories (e.g. body, agriculture, hunting, construction, zoological, familial, other).
10. I’d like to store a version of the Hebrew with no cantillation or vowel marks so that we can run searches on it more easily in the future.
11. I believe this is in the project specs (revised\_plan.md) but just in case not – we eventually want to also capture WHO is speaking (e.g. narrator, Moses, God, etc).