

Figure : Flowchart for vectorizing all included studies.

**Algorithm 1** Process for Extracting Variables from Studies

**Require:** Generate embeddings for all included studies using OpenAI’s ada v2 model with a chunk size of 500.

1: Initialize an empty list to store final results.

2: **for** study in the list of studies **do** 3: Load the embeddings for specified study.

4: **for** variable in the list of variables **do**

5: Calculate the cosine similarity to identify top 5 embedding vectors for this variable.

6: Augment these top five vectors with the original query, including the variable and the required response format.

7: Pass the augmented input to OpenAI’s ChatGPT-4o API for extraction and formatting.

8: Append the resulting output to the final results list.

9: **end for** 10: **end for**

11: Save the combined results in a Microsoft Excel sheet.