Screening Agricultural Research Papers with AI

AUTHOR
Lolita Muller

PUBLISHED
January 31, 2025

1 Screening Agricultural Research Papers with AI

This vignette demonstrates how to automate the screening of agricultural research papers using GPT-based AI models. The goal is to efficiently classify and filter research papers based on predefined inclusion/exclusion criteria, reducing manual workload and ensuring consistency in the selection process.

1.1 Step 1: Load & Sample Bibliographic Data

▸ Code

2 Step 2: AI-Based Screening with ChatGPT

To use this approach, you'll need an OpenAI account and an API key for accessing GPT models. Once you have your API key, you can customize your prompt based on your specific screening criteria.

Since API calls incur costs each time they are executed, the code is currently commented out (#) to prevent unnecessary charges. It's recommended to run the extraction process only when needed, and once you are satisfied with the results, save the output DataFrame to avoid repeated API costs.

▸ Code

▸ Code

Summary of Prompts based on inclusion criteria

Criterion	Description
Location (Latin America & Central America)	Extract country names from the abstract. If the study is conducted in Latin America or Central America, return 'Yes'. Otherwise, return 'No'. If no location is mentioned, return 'NA'.
Field Study	Determine whether the study is an **on-field experiment** (not a lab or greenhouse study). If conducted on a farm or in natural field conditions, return 'Yes'. Otherwise, return 'No'.
Primary Data	Check if the study collects **primary data** (original experimental data). If it does, return 'Yes'. If it is a review, synthesis, or modeling study, return 'No'.
Study Type	Determine whether the study is a **review, synthesis, or modeling exercise**. If it falls into any of these categories, return 'No'. If it is an experimental field study using primary data, return 'Yes'.
Focus on Agriculture	Check whether the study focuses on **agriculture**. If it explicitly investigates agricultural practices, return 'Yes'. Otherwise, return 'No'.
Target Crops/Livestock	Determine if the study focuses on **maize, coffee, beans, or cattle**. If one or more of

Target Crops/Livestock	Determine if the study focuses on **maize, coffee, beans, or cattle**. If one or more of these are studied, return 'Yes'. If the study focuses on other crops or livestock, return 'No'.
Tested Agricultural Practices	Extract the **agricultural practices tested in the field** from a predefined list (e.g., agroforestry, crop rotation, organic fertilizer). If no listed practices are tested, return 'No'.

▸ Code

Show 5 entries Search:

Abstract Note	location_LA_CA	field_study	primary_data	study_type	focu
▸ Click to expand	China	Yes	Yes	Yes	Yes
▸ Click to expand	Brazil	Yes	Yes	Yes	Yes
▸ Click to expand	China	Yes	Yes	Yes	Yes
▸ Click to expand	Brazil	Yes	Yes	Yes	Yes
▸ Click to expand	China	No	No	No	Yes

Showing 1 to 5 of 100 entries Previous 1 2 3 4 5 ... 20 Next

2.1 Accuracy

▸ Code

GPT Screening Accuracy: 81 %

Screening_Comparison	n
False Exclusion (GPT Error)	14

False Inclusion (GPT Error)	5
Match - Included	10
Match - Rejected	71

Note: Some of the false exclusions identified by GPT are actually correct. During the ACDC extraction process, we unintentionally included some papers from Africa, even though they should have been excluded based on the criteria. In these cases, GPT correctly excluded the papers based on location, while our original extraction mistakenly retained them.

2.2 Cost analysis

▸ Code

GPT Processing Cost Summary

Metric	Value
Total Cost for 100 Abstracts	$3.35
Total Requests Made	700
Cost per Abstract (All 7 Requests)	$0.0335
Cost per GPT Request	$0.00479

▸ Code

Estimated Costs for Different Dataset Sizes

Number.of.Abstracts	Total.Requests	Estimated.Cost...USD.
1e+02	700	3.35
5e+02	3500	16.75
1e+03	7000	33.50
5e+03	35000	167.50
1e+04	70000	335.00
5e+04	350000	1675.00
1e+05	700000	3350.00