Reproducibility of deep learning pipeline method information using a multi-modality approach

Vamsi Krishna Kommineni[1,2,3], Waqas Ahmed[1], Sheeba Samuel[1,4], Birgitta König-Ries[1,2,4]

[1] Heinz Nixdorf Chair for Distributed Information Systems, Friedrich Schiller University Jena, Germany
[2] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
[3] Max Planck Institute for Biogeochemistry, Jena, Germany
[4] Michael Stifel Center Jena

Introduction

- Reproducibility involves obtaining consistent results upon repeating an experiment.
- Retrieving all the information related to Deep Learning (DL) methods is challenging.
- Reproducibility variables of DL method information: data, data preprocessing steps, source code, hyperparameters, model information, post processing, evaluation metrics etc.
- Useful DL pipeline information available in source code, training curves, text, tables, figures, and images.
- Create a separate pipeline for each data modality to convert the text summaries.

- Generate queries and then infer LLM using combined text summaries to get answers.
- In this work, we leverage multi-modal information to retrieve the complete available DL method information from the target domain scholarly publications.
- With the extracted information, we will check the reproducibility status of the DL pipeline used in publications.
- The whole pipeline will be built using open-source models.
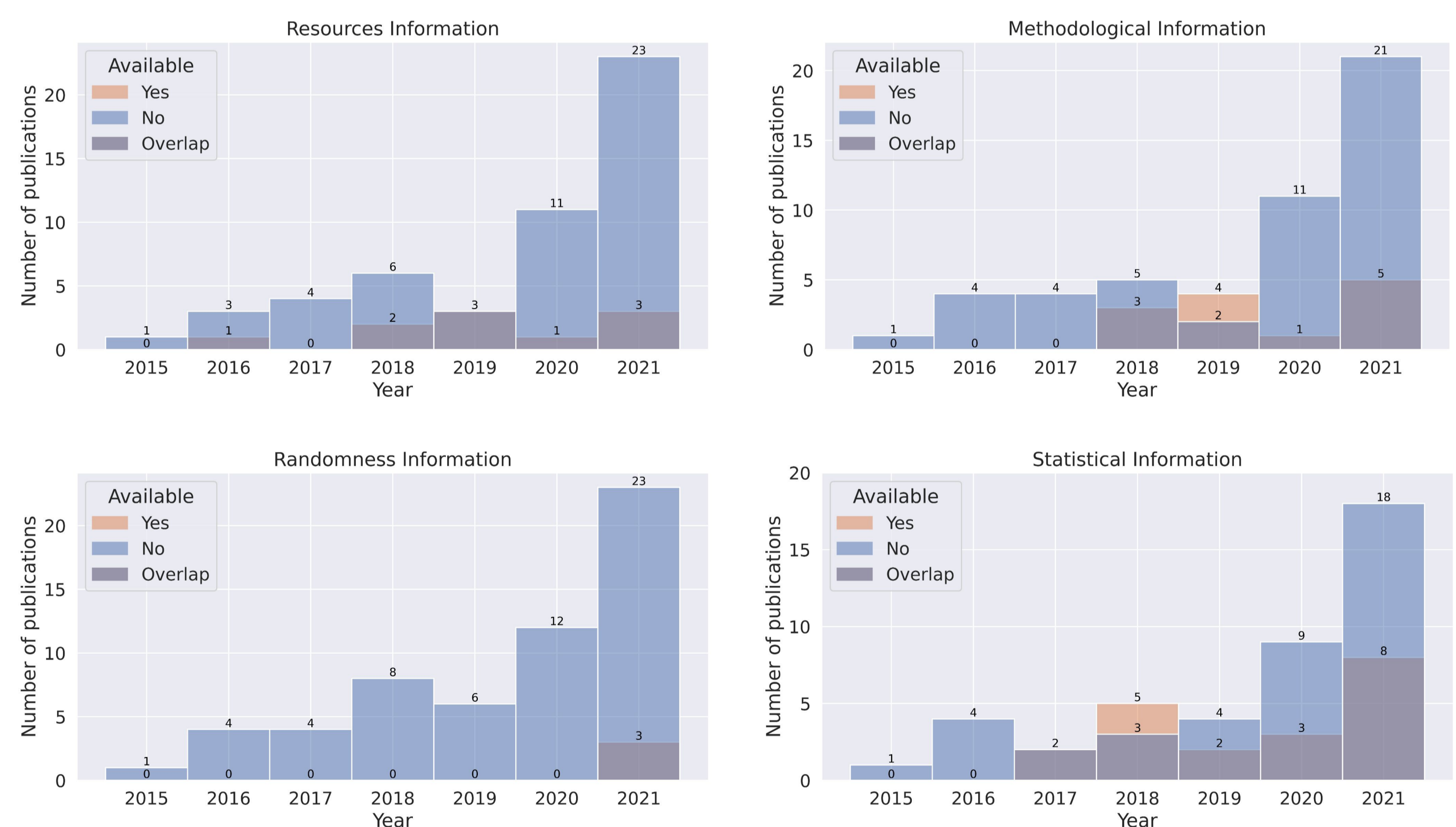
Method
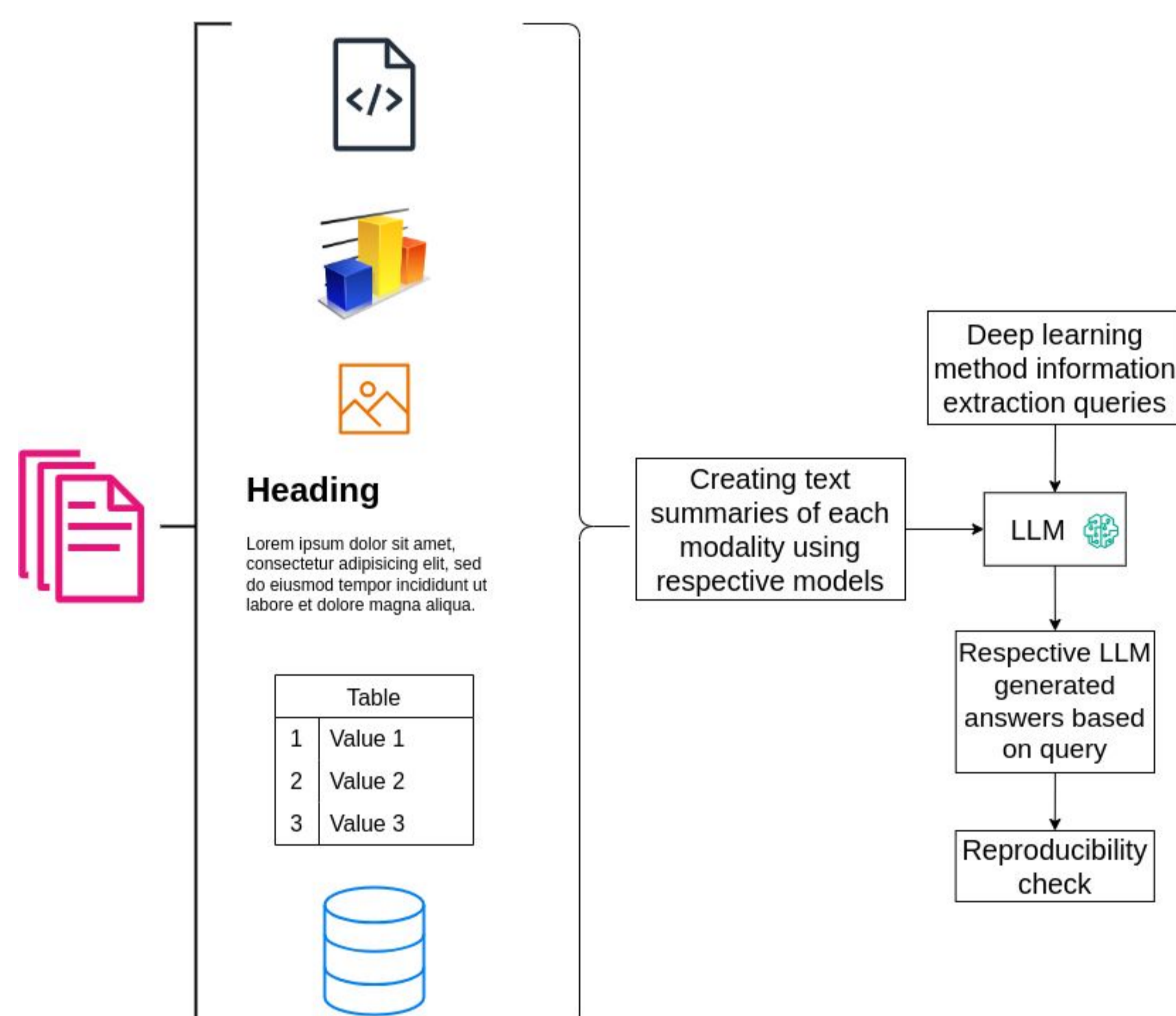
- Code processing: Use open source model like code Llama to extract the information about hyperparameters, model information, random seeds, software packages used and so on
- Graph / Figures: Summarize information within figures and graphs with models like LLaVA.
- Raw text: Employ LLM (e.g., Llama-2-70b, Mixtral-8x7B-v0.1, Mistral Large etc) to process textual data
- Tables: Summarize table information or explore all possible methods to extract data from tables into natural language text.
- Data: Summarize available datasets using LLM and utilize this information to align with dataset details in the raw text.
- Integrate different modalities: Combine various modalities to get the complete DL pipeline information.
- Evaluation: Evaluate each modality and the pipeline with integrated multi modal data results with tools like Ragas.
- Queries: Generate detailed queries related to DL pipeline method information.
- Reproducibility check: Employ combined multimodal text summaries, queries, and generated responses to determine the number of queries with positive outcomes. To further assess reproducibility, categorize queries and analyze positive responses within each group.

Challenges to solve

- Prompt engineering plays a major role in enhancing the outputs of LLMs. Hence, the proper experimentation of different versions of the prompts will be needed to understand the specific model limitations.
- With the ever-changing landscape of LLMs, selecting the available state-of-the-art model while conducting the experiments is essential.
- Explore different combinations of multi-modal models, summarize the text from other modalities, and use that to infer LLM.

Initial results

Below are the results generated for the text processing pipeline alone

Number of papers meeting criteria for the four categories 1) Resources 2) Methodological information 3) Randomness 4) Statistical consideration for selected research publications by year

- The data and the code used to extract and analyse the reproducibility information of DL methods from publications for text modality is publicly available: https://github.com/fusion-jena/Reproduce-DLmethods-Biodiv

Conclusion

- We presented our pipeline for assessing the reproducibility of deep learning methods in biodiversity research.
- A semi-automatic approach that leverages Large Language models for extracting information on reproducible variables from publications.
- A pipeline to integrate different modalities to get the complete information of the DL methods from scholarly publications.

References:

1. https://github.com/fusion-jena/Reproduce-DLmethods-Biodiv
2. Ahmed, W., Kommineni, V.K., K¨onig-ries, B., Samuel, S.:How Reproducible are the Results Gained with the Help of Deep Learning Methods in Biodiversity Research?. Biodiversity Information Science and Standards, 7. (2023)