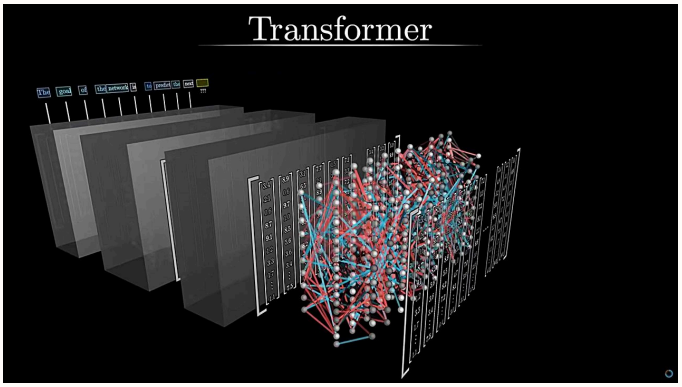
EXPLORING CHATGPT FOR DATA CLEANING AUTOMATION IN DATA SCIENCE

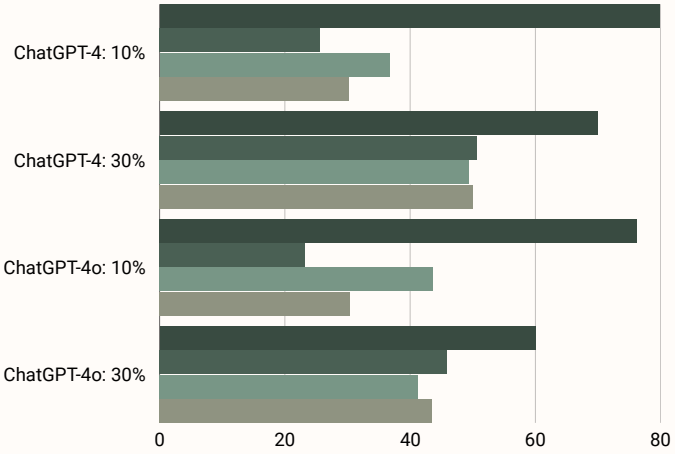


CHATGPT-4 & CHATGPT-4OMNI

Data cleaning consumes significant time for data scientists, detracting from analysis. ChatGPT-4 and its lightweight variant, ChatGPT-4 Omni are evaluated for streamlining data cleaning, using accuracy, precision, and recall as metrics. ChatGPT-4 prioritizes precision, while Omni balances efficiency and performance for everyday tasks with lower computational demands.

MODEL & ERROR-RATE IMPUTED DATASETS

- ChatGPT-4: 10%
- ChatGPT-4o: 10%
- ChatGPT-4: 30%
- ChatGPT-4o: 30%



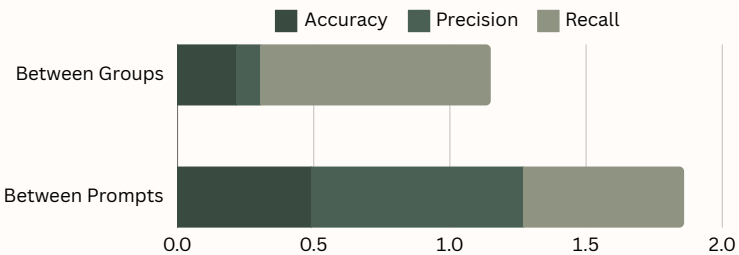
RESULTS

A confusion matrix facilitated the calculation of accuracy, precision, and recall. ChatGPT-4 and ChatGPT-4omni performed better on 10%-imputed-error datasets (accuracy: 79.8%, 76.2%) than 30% error datasets (69.9%, 60.0%). Precision was higher for 30%-imputed-error datasets (50.7%, 45.9%) versus 10% error datasets (25.6%, 23.2%). recall was consistent across all groups, averaging 42.8%.

Metrics

- Accuracy
- Precision
- Recall
- F1_score

P-VALUES



The Kruskal-Wallis test found no significant differences in metrics across groups: accuracy ($p = 0.218$), precision ($p = 0.086$), and recall ($p = 0.846$). Similarly, no significance was observed between prompts: accuracy ($p = 0.491$), precision ($p = 0.779$), and recall ($p = 0.590$), eliminating the need for post-hoc analysis.

FINDINGS

ChatGPT-4 Omni offers comparable performance to ChatGPT-4 but with lower resource demands. However, both models showed suboptimal precision, recall, and F1 scores, indicating the need for human intervention for consistency and validation of LLM outputs and effective data cleaning.

Supervisor: Prof. dr. DEPAIRE Benoît

Prepared by: ACHRAF Oussama