Using AI to Summarize and Interact with Handwritten Notes

Members: Kadija Bangura, Maanas Bysani, Ling Liu, Xianhao Xu / Supervisors: Anjali Golding, Norman Di Palo

Problem & Solution

Handwritten v.s. digital note-taking in the digital age

- Digital notes: Easy editing, concise, and fast to create.
- Handwritten: Disorganized and cluttered. Cognitive involvment boosts memory and performance.[1][2][3][4]

Our tool: Aims to bridge the gap between the best of both worlds by allowing users to:

1. Upload handwritten notes, convert to text.
2. AI searches for markers (Fig 1) of important information.
3. Interact with notes via user-friendly UI (Fig 5) to ask questions, simplify material, summarise, organise logically, and review/revise.

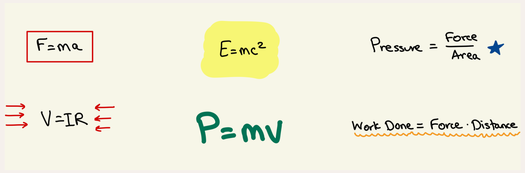

Figure 1: Examples of Markers the AI would search the image for.

Method

Technical:

- Vision models process image input into LaTeX code via relevant AI APIs - GPT4V, Claude, Gemini.
- LaTeX code is generated and processed into a PDF using the aspose.pdf library automatically once the image is uploaded on the webpage.
- User interaction with notes is facilitated through AI, with user input serving as prompts for the AI-API requests.

Model Accuracy Comparison:

- Subset of 15 characters from image containing 128 randomized characters (English and Greek alphabets, numbers, and math operators) was used to represent commonly encountered characters in university courses.
- AI models were tested with the subset, and an accuracy score was computed against a human control baseline.
- Three different subsets (Fig 2) were used and results were averaged.

Results: Model Accuracy

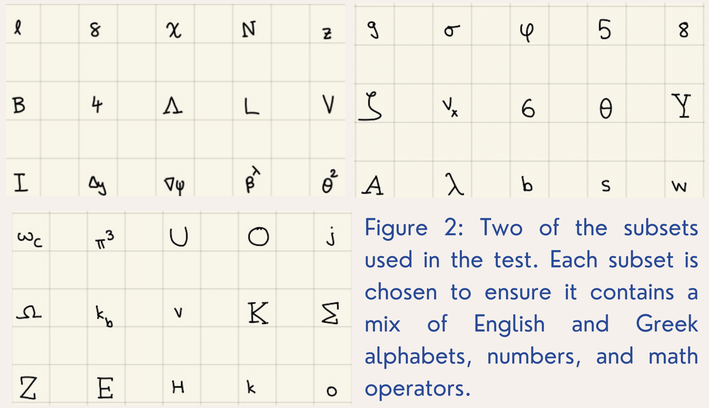

Figure 2: Two of the subsets used in the test. Each subset is chosen to ensure it contains a mix of English and Greek alphabets, numbers, and math operators.

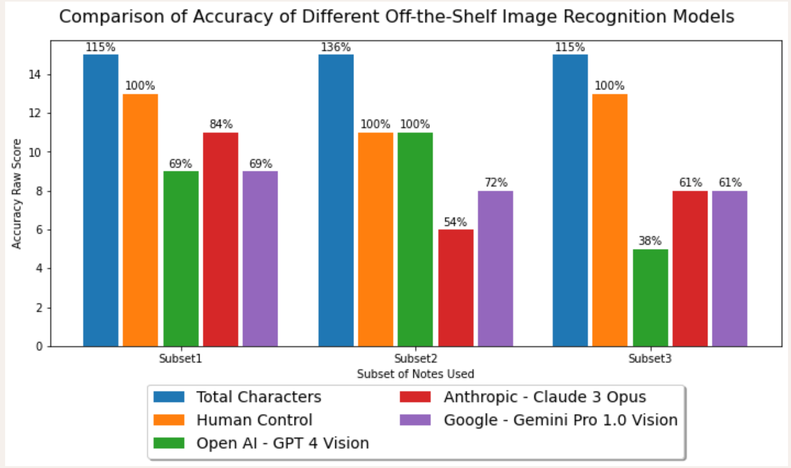

Figure 3: Accuracy scores of each Image Recognition Model, a human control and the total characters in each subset. The bars reflect the raw score; the percentages are computed against the human control baseline.

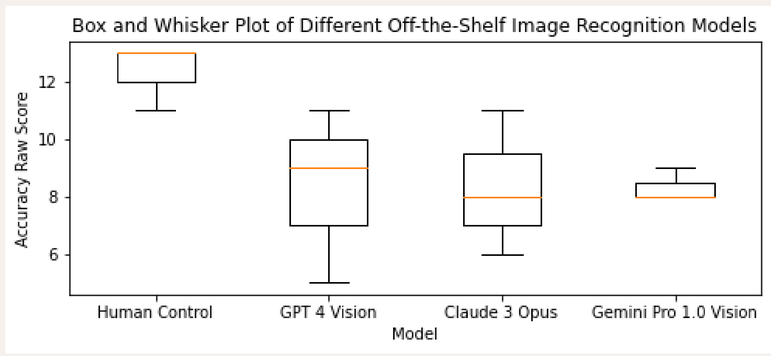

Figure 4: Box and Whisker Plot of the accuracies of each test reflecting the uncertainty involved. The seemingly large uncertainty arises due to the small sample size of three subsets.

Results: User-friendly UI

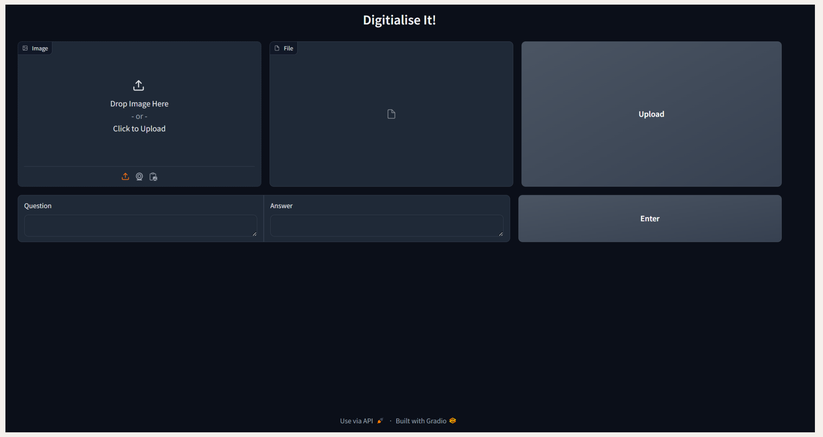

Figure 5: User-friendly UI (made using the Gradio library) where the user can upload images of their notes which returns a PDF summary (Top Row). The user can also interact with their notes through text which the AI responds to (Bottom Row).

Conclusion & Future

- GPT4 Vision, scored the highest accuracy, on average. However, Gemini Pro Vision seems to deliver the most consistent results.
- User friendly UI ready to be deployed but slight improvements to functionality and operation needed.
- If given more time, we would try different and larger characters combinations and compare accuracies.
- We would also test the user workload using surveys like the NASA TLX to understand painpoints and verify usefulness of tool.
- Next step is to train model to search for markers shown in Fig1.

References

[1]Duran, K.S. and Frederick, C.M. (2013). Information Comprehension: Handwritten vs. Typed Notes. Undergraduate Research Journal for the Human Sciences, [online] 12(1). Available at: https://publications.kon.org/urc/v12/duran.html.Year of Publication

[2]Corrigendum: The Pen Is Mightier Than the Keyboard: Advantages of Longhand Over Laptop Note Taking. (2018). Psychological Science, 29(9), 1565–1568. https://doi.org/10.1177/0956797618781773

[3]Olive, T. and Barbier, M.-L. (2017). Processing Time and Cognitive Effort of Longhand Note Taking When Reading and Summarizing a Structured or Linear Text. Written Communication, 34(2), pp.224–246. doi:https://doi.org/10.1177/0741088317699898.

[4] Voyer, D., Ronis, S.T. and Byers, N. (2022). The effect of notetaking method on academic performance: A systematic review and meta-analysis. Contemporary Educational Psychology, 68(1), p.102025. doi:https://doi.org/10.1016/j.cedpsych.2021.102025.