Can LLMs outperform classical OCR/HTR pipelines?

Analysis of LLMs' capabilities and limitations in transcribing old handwritten historical records.

Seorin Kim [1] Julien Baudru [2] Hugues Bersini [2] Vincent Ginis [1]

[1]Vrije Universiteit Brussel (VUB) [2]Université Libre de Bruxelles (ULB)

Data

- 20 high-resolution scans of *Déclarations de succession et de mutation par décès*.

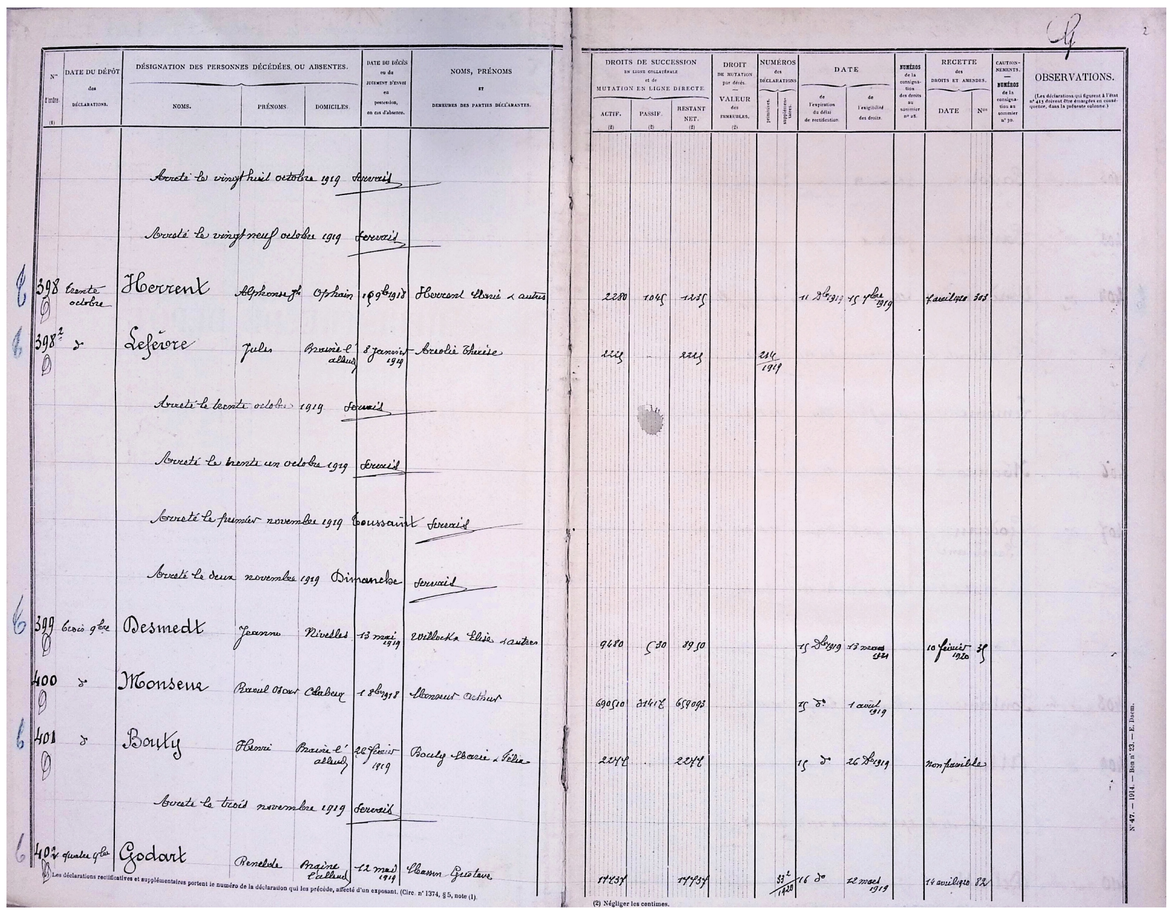

Figure 1. Example of *Déclarations de succession et de mutation par décès.*

Workflow

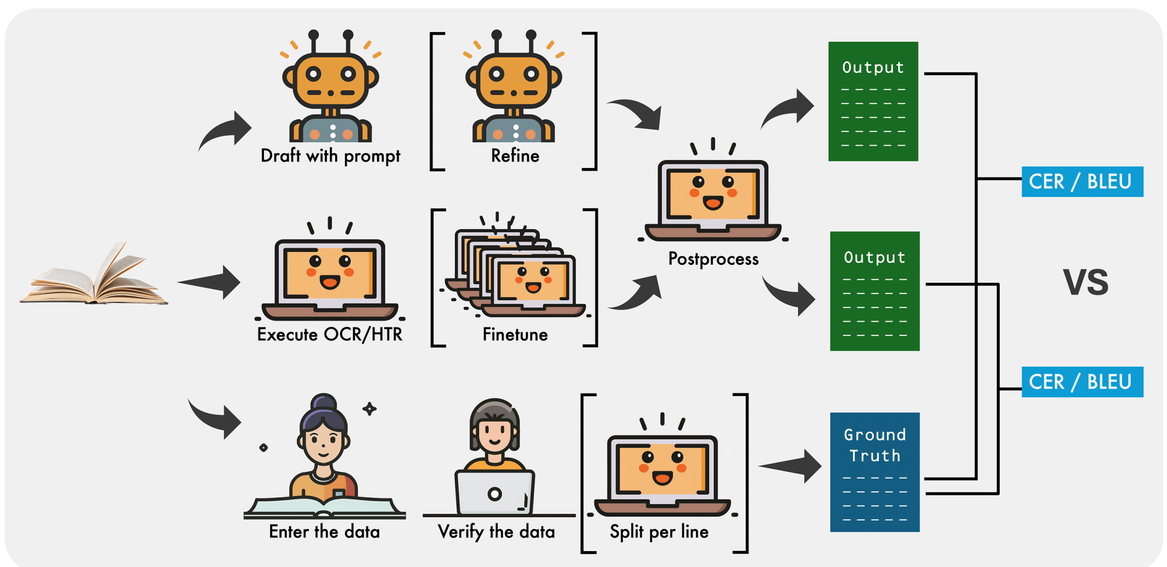

Figure 2. Workflow for the LLMs, OCR tools, and manual transcription.

1. Create the ground truth data.
2. Process scanned pages by LLMs with 4 different prompts (simple, complex prompt, prompt with one example, and two examples) and methods (one/two-shots, and refinement).
3. Process the pages by 4 OCR/HTR systems without fine-tuning and with fine-tuning for trOCR.
4. Postprocessing the outputs (e.g., removing separators and delimiters).

[1] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318, 2002.

Results

1. Character Error Rate (**CER**): $CER = \frac{(S+D+I)}{(S+D+C)}$ S =Substitutions, D =Deletions, I =Insertions, C = Correct Characters
2. Bilingual Evaluation Understudy (**BLEU**) [1] : "An algorithm for evaluating the quality of text which has been machine-translated from one natural language to another.". The score lies between 0 and 1 (= highest similarity).

Compared Models:

- **LLMs**: *ChatGPT-4.0* and *Claude 3.5 Sonnet*
- **OCR/HTR**: *EasyOCR*, *Pytesseract*, *KerasOCR*, and *trOCR*

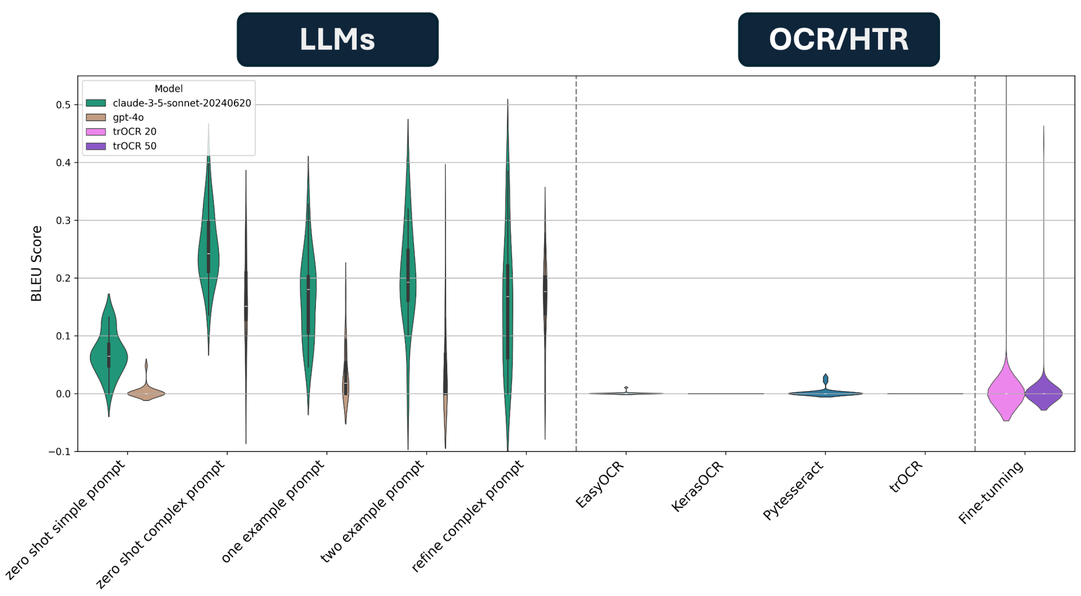

Figure 3. **BLEU** metric comparison for each method.

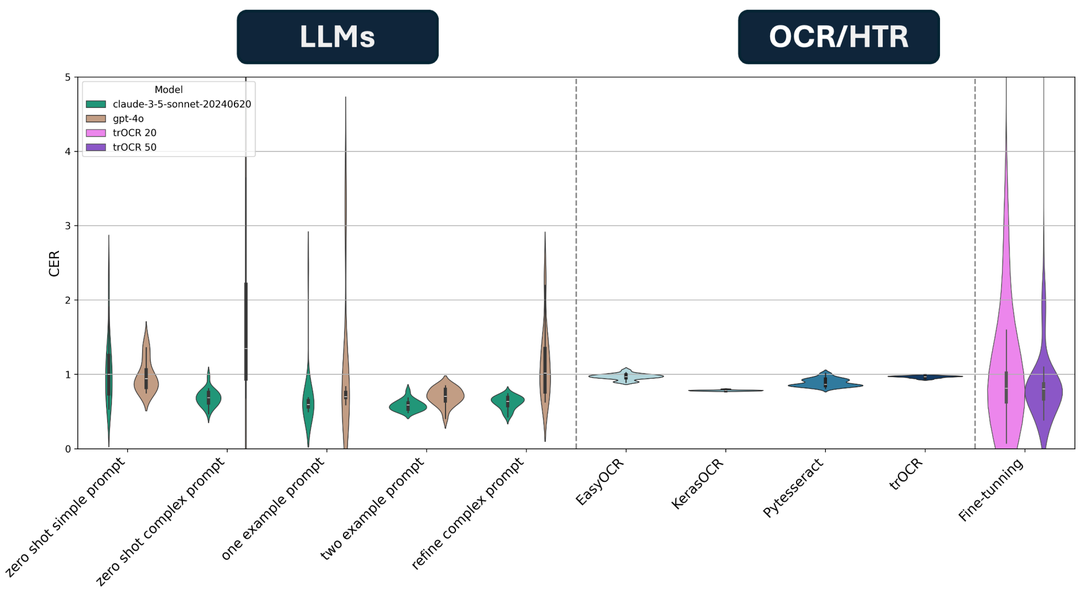

Figure 4. **CER** metric comparison for each method.

Outputs Comparison

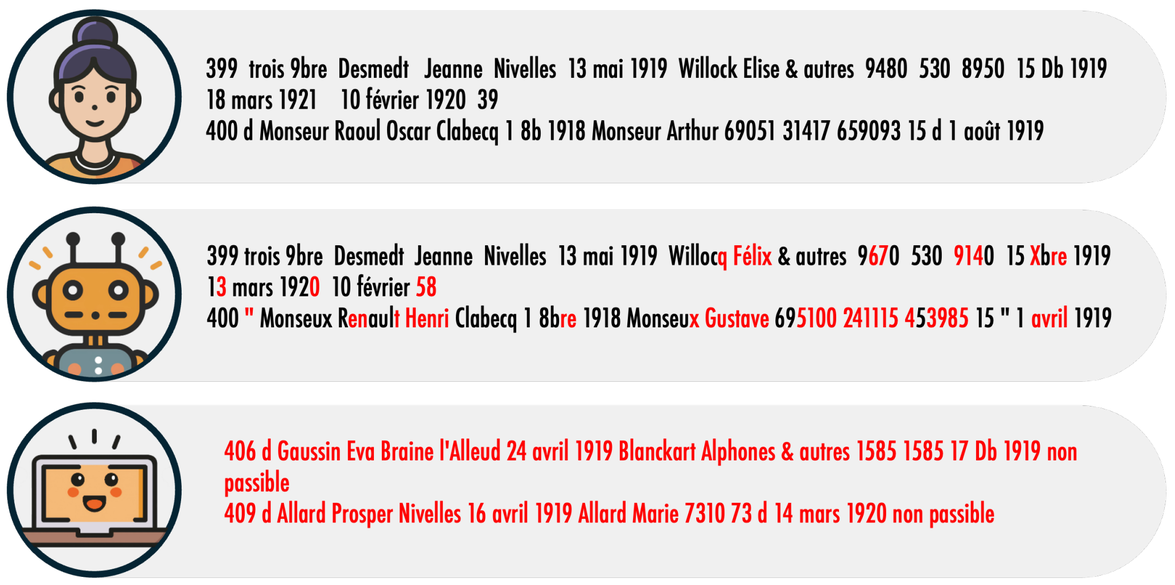

Figure 5. 9th and 10th lines of Figure 1 transcribed by the authors, Claude 3.5 Sonnet with two examples, and TrOCR fine-tuned with 20% of the data.

seorin.kim@vub.be & julien.baudru@ulb.be