

Otto-Friedrich-Universität Bamberg



LLMs vs. Traditional Models: A Task-Based Analysis of NLP Use Cases



Robin Jegan, M.Sc.
University of Bamberg
Chair of Media Informatics
robin.jegan@uni-bamberg.de

Context



- Since 2019 at University of Bamberg
 - PhD Student
 - Research assistant at Chair of Media Informatics
- DH projects
- Since 2021: Teaching
 - Supervised 13 Student's thesis on NLP use-cases
- Research
 - Domain-specific use cases and trends in NLP
 - Applicability of traditional approaches



NLP in Recent Years: My Perspective



- Overarching scenario – text segmentation [3]
 - Video transcripts of a lecture on software engineering
- Can traditional models keep up?
- Motivation: Limitations of current LLMs
 - Efficiency
 - Reproducibility & Explainability
- For text segmentation
 - Previous study using GPT-3.5-turbo [6]:
 - Unreliable outputs
 - Ignored instructions given in the prompt

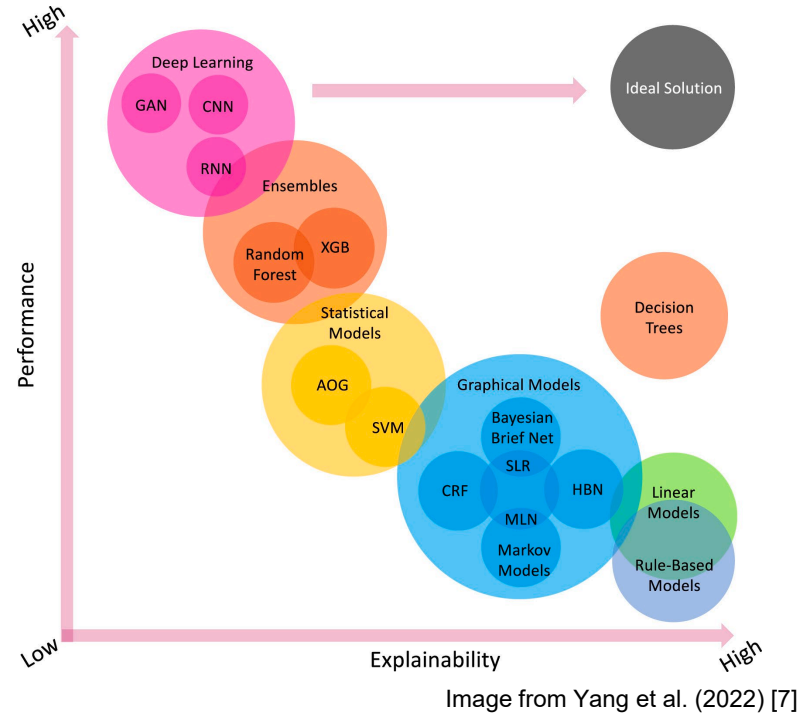


- Examples
 - Rule-based techniques & decision trees
 - Extractive summarization systems
 - SVM, Random Forest & Naive Bayes Classifiers
 - And many more: Jegan et al. (2025) [4]
- For text segmentation
 - TextTiling by Hearst (1997) [2]
 - Lexical co-occurrence and distribution
 - Detect segments through subtopic shifts

Difficulties for LLMs



- Explainability
 - Intrinsic vs. Post-hoc methods
- Reproducibility
 - Consistent output vs. Constrained creativity for LLMs
- Lacking metrics
 - Limitations of purely quantitative scoring
- For text segmentation
 - No gold standard available
 - Complex evaluation of transcripts (Whisper) for segmentation task



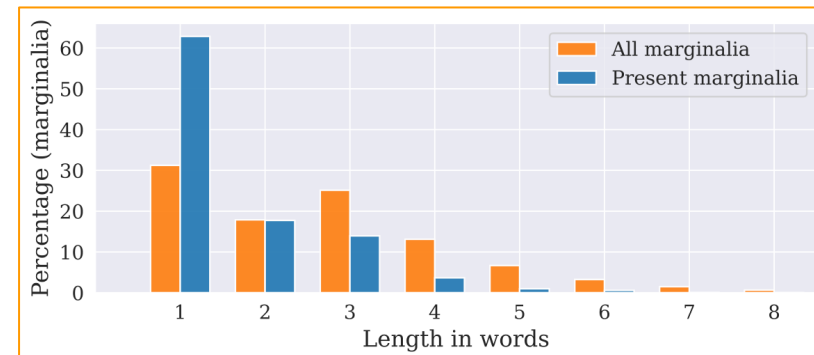
Experiments



- Information extraction on company database
 - Extraction of legal form of an organization/firm
 - GmbH, AG, GbR, etc.
- Text classification on insurance data
 - Support Vector Machines for small-data domains
 - Lower performance needs with comparable quality
- Keyphrase extraction for marginalia creation [5]
 - GPT-4o and mT5 perform best
 - TF-IDF (extractive) as baseline trailing only slightly behind

Regex:

```
( ^ | ) b g b ( | $ )
```

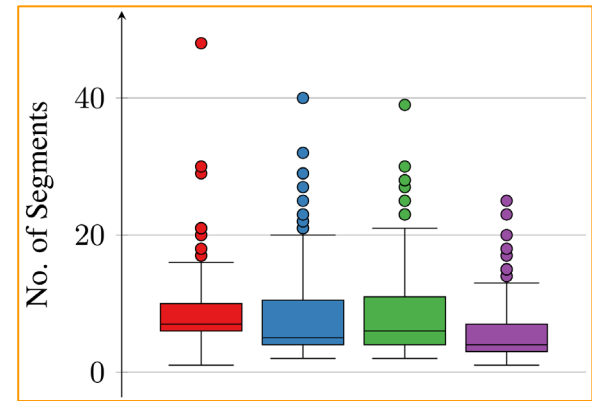


Results for Text Segmentation [3]



- Quantitative
 - Efficiency for sequential processing
 - Number of segments
- Qualitative
 - Difficulty in evaluating the quality of produced segments
 - Results from both models comparable

Model	Energy Consumption in Wh	Execution Time in sec
GPT-4o	51.6 - 516 [1, 8]	6127
TextTiling	0.085 - 0.47	9 - 16



Red: GPT-4o; others: TextTiling with various configurations
Image from Jegan et al. (2025) [3]



- LLMs outperform traditional techniques on „open“ tasks
 - In summarizing or transforming tasks
 - However: Hallucinations and reproducibility issues
- But: Traditional techniques still relevant and promising for
 - Constrained tasks
 - Clearly defined or marked use-cases
 - Small-data domains



Thanks!

Open to research and
industry opportunities!
mail@rjegan.de

References



- [1] **De Vries**, Alex. "The growing energy footprint of artificial intelligence." *Joule* 7.10 (2023): 2191-2194.
- [2] **Hearst**, Marti A. "Text tiling: Segmenting text into multi-paragraph subtopic passages." *Computational linguistics* 23.1 (1997): 33-64.
- [3] **Jegan**, Robin, and Andreas Henrich. "Contrasting Traditional Models and LLMs: An Evaluation Based on Text Segmentation." *Proceedings of the 21st Conference on Natural Language Processing (KONVENS 2025): Workshops*. 2025.
- [4] **Jegan**, Robin, and Andreas Henrich. "A Structured Literature Review on Traditional Approaches in Current Natural Language Processing." *arXiv preprint arXiv:2505.12970* (2025).
- [5] **Lang**, Aaron, Robin Jegan, and Andreas Henrich. "Automatic Creation of Marginalia." *Proceedings of the 21st Conference on Natural Language Processing (KONVENS 2025): Long and Short Papers*. 2025.
- [6] **Lehmann**, Alexander, and Dieter Landes. "Extracting Metadata from Learning Videos for Ontology-Based Recommender Systems Using Whisper & GPT." 2024 IEEE Global Engineering Education Conference (EDUCON). IEEE, 2024.
- [7] **Yang**, Guang, Qinghao Ye, and Jun Xia. "Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond." *Information Fusion* 77 (2022): 29-52.
- [8] **You**, Hosh. "How much energy does ChatGPT use?". *epoch.ai*. <https://epoch.ai/gradient-updates/how-much-energy-does-chatgpt-use>