

Direct and Indirect Annotation with Generative AI: A Case Study into Finding Animals and Plants in Historical Text

Arjan van Dalmsen^{1,*}, Folgert Karsdorp², Ayoub Bagheri³, Dieuwertje Mentink¹, Thirza van Engelen¹ and Els Stronks¹

¹Department of Language, Literature and Communication, Utrecht University, Trans 10, Utrecht, 3512 JK, The Netherlands

²KNAW Meertens Instituut, Oudezijds Achterburgwal 185, 1012 DK Amsterdam, The Netherlands

³Department of Methods and Statistics, Utrecht University, Padualaan 14, 3584 CH, Utrecht, The Netherlands

Abstract

This study explores the use of generative AI (GenAI) for annotation in the humanities, comparing direct and indirect annotation approaches with human annotations. Direct annotation involves using GenAI to annotate the entire corpus, while indirect annotation uses GenAI to create training data for a specialized model. The research investigates zero-shot and few-shot methods for direct annotation, alongside an indirect approach incorporating active learning, few-shotting, and k-NN example retrieval. The task focuses on identifying words (also referred to as entities) related to plants and animals in Early Modern Dutch texts. Results show that indirect annotation outperforms zero-shot direct annotation in mimicking human annotations. However, with just a few examples, direct annotation catches up, achieving similar performance to indirect annotation. Analysis of confusion matrices reveals that GenAI annotators make similar types of mistakes, such as confusing parts and products or failing to identify entities, which are broader than those made by humans. Manual error analysis indicates that each annotation method (human, direct, and indirect) has some unique errors. Given the limited scale of this study, it is worthwhile to further explore the relative affordances of direct and indirect GenAI annotation methods.

Keywords

large language models, natural language processing, historical text, token classification, environmental humanities

1. Introduction

The introduction of advanced generative AI (GenAI) models has sparked interest among humanities scholars in leveraging these tools to extract structured information from texts [3, 23, 2, 12, 25, 18, 19, 4, 5]. So far, the use of GenAI in the humanities has primarily involved “direct

CHR 2024: Computational Humanities Research Conference, December 4–6, 2024, Aarhus, Denmark

*Corresponding author.

✉ j.a.vandalfsen@uu.nl (A. v. Dalmsen); folgert.karsdorp@meertens.knaw.nl (F. Karsdorp); a.bagheri@uu.nl (A. Bagheri); d.l.mentink@students.uu.nl (D. Mentink); t.w.e.vanengelen@students.uu.nl (T. v. Engelen); e.stronks@uu.nl (E. Stronks)

🌐 <https://www.karsdorp.io/> (F. Karsdorp); <https://ayoubbagheri.nl/> (A. Bagheri)

🆔 0000-0002-4209-4063 (A. v. Dalmsen); 0000-0002-5958-0551 (F. Karsdorp); 0000-0001-6366-2173 (A. Bagheri); 0000-0001-9741-7264 (E. Stronks)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

annotation”, where GenAI analyzes a corpus without further interference. This approach has shown promise, potentially “supercharging the humanities” [11].

Researchers in Natural Language Processing (NLP) have proposed an alternative “indirect annotation” framework. This two-step process involves GenAI generating training data, which is then used to train a specialized model. This approach offers potential cost and performance advantages over direct annotation [31, 32]. However, indirect annotation’s effectiveness has primarily been demonstrated on languages well-represented in GenAI training data, raising questions about its applicability to texts from smaller languages or historical variants often encountered in humanities research.

As a first exploration into its usability in the humanities, our study tests GenAI as an indirect annotator for nature-entities in historical Dutch texts. We employ the LLMaAA (Large Language Models as Active Annotators) framework [32], which combines few-shotting, k-Nearest Neighbors (k-NN) example retrieval, and active learning. Our research compares the performance of indirect annotation, i.e. an LLMaAA-derived model, against both human annotations and direct annotation with GenAI. We find that our proposed method of indirect GenAI annotation performs better than fully-unsupervised direct GenAI annotation. However, we also find that providing direct annotation with demonstrations (i.e., examples of annotations) results in similar performance. Moreover, our study reveals that humans, direct GenAI annotators, and indirect GenAI annotators each have unique weaknesses and strengths.

This study is structured as follows: We first examine the broader context of using GenAI for annotation in humanities research. We then provide an overview of current research on direct and indirect annotators in NLP. Subsequently, we introduce our specific use-case: identifying animals and plants in historical texts. Finally, we detail our methodology for comparing the performances of human annotation, direct GenAI annotation, and indirect GenAI annotation.

2. Related Work

2.1. GenAI annotations in humanities

In the humanities, research on GenAI annotation has primarily focused on direct annotation experiments. Studies have compared GenAI methods with traditional approaches and human annotators across various tasks, including sentiment analysis [2, 25, 5], topic detection [18], and text classification [19]. Findings generally suggest that while GenAI often outperforms dictionary-based methods, it typically falls short of specialized models. However, Karjus [12] reports human-level annotations by GenAI across diverse tasks and languages, proposing a machine-assisted mixed methods approach. These studies underscore the potential of GenAI in humanities research, while also highlighting the need to explore both direct and indirect annotation approaches to fully leverage its capabilities.

2.2. GenAI as direct annotators in NLP

Direct GenAI annotation involves prompting GenAI to annotate a dataset for immediate use. Studies assessing this approach have found that while GenAI generally lags behind state-of-the-art models [9, 16, 24, 33], it often equals or outperforms crowd-workers [30, 33, 10]. Challenges

in direct GenAI annotation include difficulties with long-tail target types, irrelevant context, and specific tasks like sequence tagging [16, 24]. These limitations have led to the exploration of indirect annotation methods, which aim to address these shortcomings by integrating GenAI in a more targeted manner.

2.3. GenAI as indirect annotator in NLP

In what we coin the indirect GenAI annotation framework, GenAI is not employed to perform the entire annotation task on a given dataset. Instead, it is used to annotate a specific subset of the dataset. This annotated subset is then subjected to further fine tuning by another model, such as a BERT model.

Wang et al. [31] found models trained on GenAI-annotated data equal to human-annotated models and outperforming direct use of GenAI. Ding et al. [7] largely echo this but also highlight a practical problem when it comes to textual analysis: GenAI is good at finding entities, but oftentimes struggles with defining the boundaries of these entities. Li et al. [20] propose a CoAnnotating framework, in which GenAI output-uncertainty is measured and annotations with the highest uncertainty (i.e., a lack of result robustness when confronted with small prompt perturbations) are sent to a human annotator. While they report promising results, there is the disadvantage of higher costs. With Large Language Models as Active Annotators (LLMaAA) by Zhang et al. [32], the idea is to use active annotation to make the downstream specific model better. It includes:

- Few-shotting: putting exemplary annotations in the prompt for the GenAI (this helps GenAI to annotate [21]);
- k-NN example retrieval: sequence embeddings of the text to annotate and the examples are used to select the examples closest to the new text for few-shotting;
- Training cycles: doing step-by-step training, where first a specific model is trained, new data is annotated by the GenAI, and the specific model is trained again;
- Active learning: selecting examples for indirect annotation for which the current model struggle;
- Automatic reweighting: assigning learnable weights to the annotated training samples [27] (this makes it possible to reduce the impact of noisy labeling by GenAI).

The authors test their method for NER and Information Extraction (modern Chinese and modern English) and find that the resulting model strongly outperforms zero-shot direct annotation with GenAI. In comparison to few-shot GenAI annotation (with k-NN optimized examples), LLMaAA shows a marginal performance advantage, besides the obvious advantages when it comes to robustness, costs, and speed, making it promising for humanities research.

2.4. Plants and animals

In this study, we research the detection of plants and animals in historical texts, which can be seen as the traditional NLP task of token (in sequence) classification or NER. Roughly starting with publications as *Man and the Natural World* [29] and *The Animal Estate* [28], humanities' scholarly interest in nature has skyrocketed. This is only natural, considering a widely shared

sense of humanity being in environmental, ecological, and climate crises. For cultural historians, the main question has been how humans observed, interpreted, and represented and thus perceived nature [22]. Research on this topic has virtually exclusively been done qualitatively. Although this approach provides illuminative insights, a qualitative approach will always work with a relatively narrow scope because of the sheer size of the historical record, thus resulting in a less precise large-scale overview of the studied phenomenon. It has, therefore, great potential to complement qualitative studies with quantitative research.

3. Methods

In this section, we describe the methodology employed to compare direct and indirect GenAI annotation strategies for identifying plants and animals in historical Dutch texts. First we explain the data parsing used to prepare our dataset. Following this, we describe the annotation procedure undertaken by human annotators. Then, we address the dataset creation. Subsequently, we describe the token classification. After this, we describe our prompts. Then, we document the training process of the indirect annotation models. Finally, we outline the ways in which we compare the annotation approaches.

3.1. Data parsing

Our study used the *Digitale Bibliotheek voor de Nederlandse Letteren* (DBNL) [6], comprising about 1,500 diverse Dutch texts. After preprocessing, the corpus yielded approximately 7 million unique sentences [17, 8].

3.2. Annotation procedure

Two texts from the 1750s were selected for manual annotation by two Early Modern Dutch literature experts. 200 sentences, parsed to have a minimal length of 10 and a maximal length of 100 words, were annotated using the INCEpTION tool [15] (cf. Fig. 1), following iteratively developed guidelines (Appendix A). The annotation schema tagged entities on three levels: Category (Plants/Animals), Type (Organism/Part/Product/Collective), and Usage (Literal/Symbolical/Petrified). For example, in “The bear grabbed an apple with its claw”, “bear” would be tagged as Animals-Organisms-Literal.

3.3. Task description

The annotated sentences were split into demonstration, validation, and test sets. We conceptualized the detection of animals and plants as a token classification task. Prompts for both direct and indirect annotators included the annotation schema, with technical details omitted to improve performance (full prompts in Appendix B, model settings in Appendix C).

3.4. Training indirect annotation models

For indirect annotation, we adapted the LLMaAA framework by Zhang et al. [32], integrating it with Huggingface and OpenAI ecosystems. We used GysBERT [1] for historical Dutch,

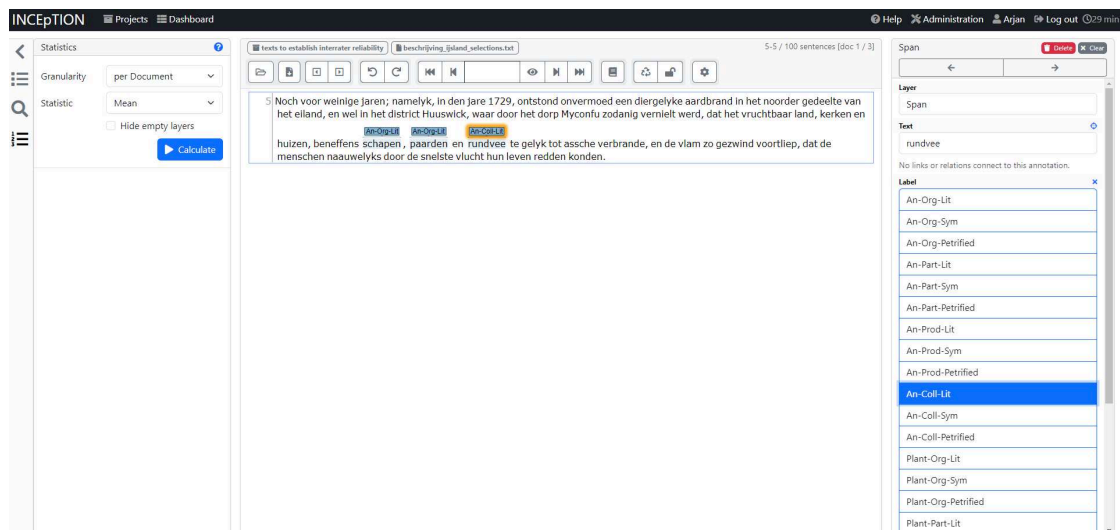


Figure 1: The annotation interface of the INCEpTION tool, which was used for conducting the human annotations.

applied k-NN few-shot selection (with the paraphrase-multilingual-mpnet-base-v2 sentence transformer [26]), and confidence-based active learning. Automatic reweighting was not included. GPT4o served as the LLM backbone. To address the scarcity of plant and animal entities, we employed a pre-filtering strategy using GPT-3.5. The specialized model underwent 10 training rounds with 10 epochs each, adding 50 example sentences per round (25 pre-filtered sentences + 25 sentences with lowest confidence). This process was repeated for two datasets, using five distinct random seeds, resulting in 10 indirect models.

It is important to point out that the indirect annotators have not seen any of the human annotations in their training regime. However, the 500 sentences they are trained on are annotated with the help of the human-annotated demonstration set and the performance of the model is determined by its score on a human-annotated validation set.

3.5. Comparing annotation strategies

All of the analyses below were done on the annotations of the various strategies on the held-out test set. Note that the sentences in this set came from the same documents as the demonstration and validation.

To assess the inter-annotator agreement among human annotators, direct GenAI, and indirect GenAI approaches, we conducted inter-annotator agreement analysis, confusion matrix analyses, and performed a manual error analysis. We compare each human annotator’s results against each automatic system’s output:

1. Human Annotations

- Human1: Annotations from the first human annotator
- Human2: Annotations from the second human annotator

2. Direct GenAI Annotations

- Direct zero-shot: Zero-shot direct annotation (without examples)
- Direct few-shot 1: Few-shot direct annotation using examples from Human1
- Direct few-shot 2: few-shot direct annotation using examples from Human2

3. Indirect GenAI Annotations

- Indirect1: Indirect annotation using examples from Human1
- Indirect2: Indirect annotation using examples from Human2

For the inter-annotator agreement, positives-only weighted F1 is used as a metric. The positives-only weighted F1 is the weighted average of all harmonious means of precision and recall of labeled entities. Thus, words that were not labeled as referring to a plants or animals related word, which is by far the most common category in this situation, are disregarded. For all GenAI annotations (direct and indirect), predictions were done five times. The average and standard deviation of the inter-annotator agreements were calculated from these iterations. The observed low variance across models and approaches suggests that these results are likely stable, despite the relatively small number of iterations.

The confusion matrices were made by comparing the human annotations to single instances of the other strategies. It should be noted that *Human1* had labeled one example as "None-label" (a token was tagged but no label was chosen), which was removed later for the process of making the confusion matrices. In addition to the confusion matrix analysis, we performed a manual error analysis on the annotations for the held-out test set.

4. Results

4.1. Inter-annotator agreement

Table 1 shows the inter-annotator agreement using positives-only weighted F1 as a metric. Key findings include:

1. Annotators of the same type resemble each other best (e.g., Human1 is closest to Human2).
2. Zero-shot direct annotation demonstrates lower internal coherence (F1 = 0.74) compared to few-shot direct (F1 = 0.9 and 0.88) and indirect annotation (F1 = 0.81 and 0.86).
3. Direct zero-shot annotation consistently underperforms, while annotations from the other human annotator achieve the highest agreement.
4. Few-shot direct and indirect annotations perform similarly, falling between zero-shot and human performance.
5. GenAI models don't simply mimic the specific human annotator they were trained on, but generalize from the examples.

These results suggest that indirect annotation and few-shot direct annotation are more reliable methods for replicating human-like annotations compared to zero-shot approaches. The choice between these methods may depend on factors beyond performance, such as ease of implementation or specific task requirements.

Table 1

Inter-annotator agreement (positives-only weighted F1): The first two columns of the table display the annotations treated as a reference point, the so-called “gold labels”. The corresponding rows indicate the level of agreement of the other strategies with these gold labels.

Method	human1	human2	direct zero-shot	direct few-shot1	direct few-shot2	indirect1	indirect2
human1	1.0000						
human2	0.8332	1.0000					
direct zero-shot	0.5456 ± 0.0629	0.5206 ± 0.0856	0.7352 ± 0.1572				
direct few-shot1	0.5939 ± 0.0256	0.5842 ± 0.0341	0.6258 ± 0.0869	0.9039 ± 0.0523			
direct few-shot2	0.5738 ± 0.0212	0.6002 ± 0.0249	0.6090 ± 0.0756	0.8323 ± 0.0290	0.8822 ± 0.0650		
indirect1	0.5481 ± 0.0524	0.5489 ± 0.0330	0.4828 ± 0.0718	0.6927 ± 0.0419	0.6805 ± 0.0428	0.8099 ± 0.1163	
indirect2	0.5753 ± 0.0277	0.5785 ± 0.0139	0.5099 ± 0.0746	0.7218 ± 0.0328	0.7149 ± 0.0400	0.7976 ± 0.0589	0.8555 ± 0.0773

4.2. Confusion matrices

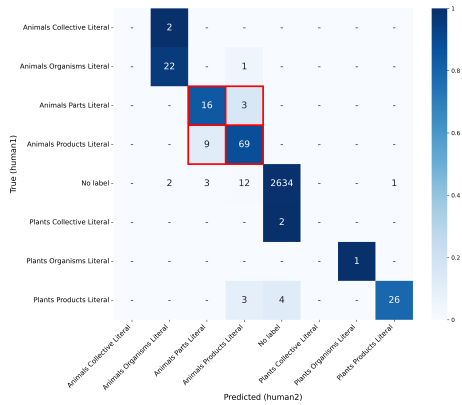
While F1 scores provide an overall measure of performance, they do not offer insights into how well the annotation methods perform for *individual* labels. To gain a more nuanced understanding of the agreement (or lack thereof) between the annotations, we turn to confusion matrices (Fig. 2), which provide insights into individual label performance across annotation methods:

1. Human2’s annotations most closely align with Human1’s.
2. Zero-shot direct annotation shows low recall, with many entities remaining unlabeled.
3. All GenAI annotators struggle with both precision and recall:
 - Precision errors: confusion between “Animals Parts Literal” and “Animals Products Literal”.
 - Recall errors: suggesting “No Label” for entities labeled by Human1.
4. GenAI methods, including indirect approach, produce labels not found by human annotators, demonstrating higher label diversity.

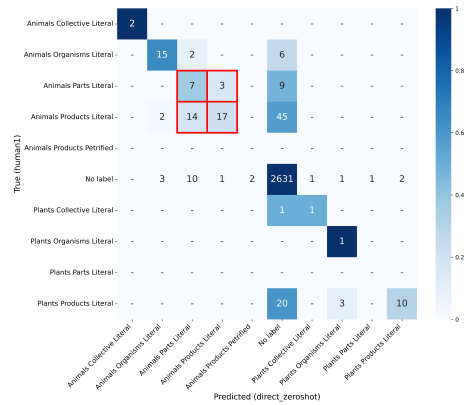
These patterns highlight the strengths and weaknesses of each annotation method, emphasizing the need for careful selection and potential combination of approaches in annotation tasks.

4.3. Manual Error Analysis

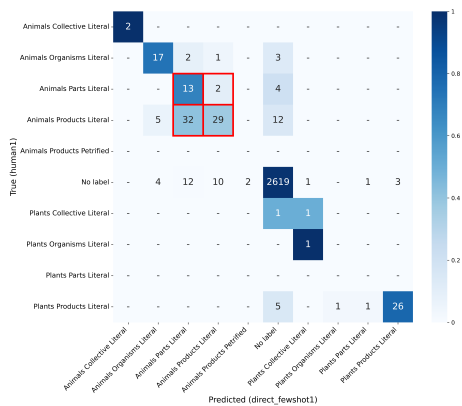
Our manual examination of the annotations (Appendix D) reveals distinct error patterns across different annotation strategies. First, human annotators occasionally overlook words in a sentence. For instance, in sentence 10, Human1 tagged the word *vleesch* (meat) only twice out of its three occurrences. Such errors likely stem from simple oversight rather than misunderstanding (although misunderstandings occur too). Second, few-shot direct annotation strategies sometimes struggle with entity aggregation. A notable example is sentence 41, where



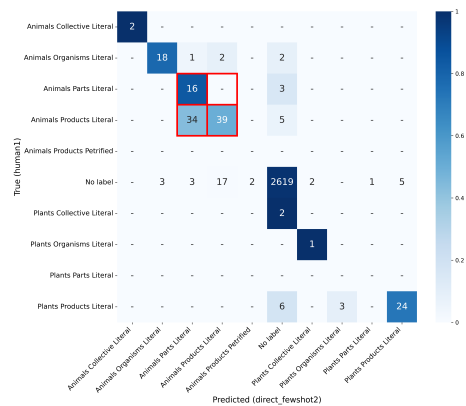
(a) Human1 vs. Human2



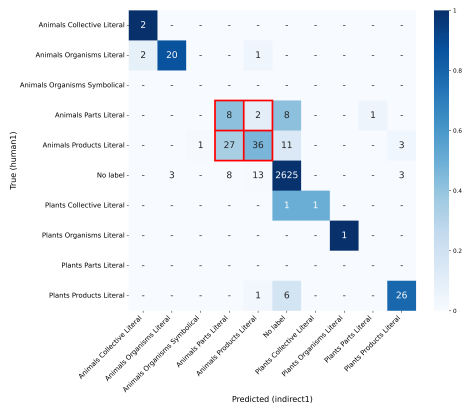
(b) Human1 vs. Direct Zero-Shot



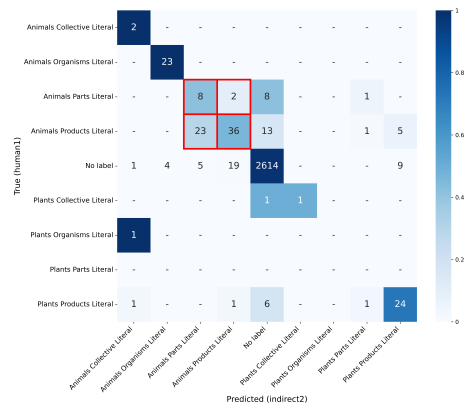
(c) Human1 vs. Direct Few-Shot1



(d) Human1 vs. Direct Few-Shot2



(e) Human1 vs. Indirect1



(f) Human1 vs. Indirect2

Figure 2: Confusion matrices comparing Human1 annotations with those from Human2 and different GenAI strategies. The red box indicates animals parts and animal products, which often proves to be a hard category for the annotators to decide on. Although all GenAI annotations were done five fold, only one of these instances is used for the confusion matrix.

“nek van het Varken” (neck of the pig) is incorrectly labeled as a single entity, instead of recognizing “neck” and “pig” as separate entities (with distinct labels “Animals Products Literal” and “Animals Organisms Literal”, respectively). Importantly, these errors are incidental, not systematic, suggesting they are unlikely to be consistently repeated. Third and finally, while indirect annotation models are (once trained) deterministic, and therefore not susceptible to *incidental* mistakes, they can produce counter-intuitive systematic errors. A revealing example is sentence 31, where “salt” is misclassified as “Plant Product Literal”. This error likely stems from the proximity of salt to spices like pepper and nutmeg in the transformer model’s vector space.

5. Discussion

This study compared direct and indirect annotation with GenAI in a humanities context, focusing on identifying plant and animal-related words in Early Modern Dutch texts. While we studied a specific case, we believe that this method can be used for a wide range of applications. Our findings reveal both the potential and limitations of various annotation strategies for historical humanities studies.

Indirect annotation demonstrates clear advantages over fully-unsupervised zero-shot direct annotation, particularly in terms of recall. However, few-shot direct annotation achieves comparable performance to indirect annotation, suggesting that both approaches have merit in different contexts. Based on these results, we advise against using zero-shot direct annotations for historical humanities research. Its significantly lower recall compared to the alternatives means that many relevant entities are likely to be missed, potentially skewing research outcomes. The choice between few-shot direct annotation and indirect annotation is less clear-cut, as both display similar F1 scores. Here, time, cost, and technical considerations should be considered.

The unique error patterns suggest two important points. First, it is crucial to investigate shortcomings of chosen methods on a micro-level to be aware of specific pitfalls. Second, there’s potential for stacking annotation methods: human, direct, and indirect annotation can be applied to the same texts, after which points of contention can be analyzed. In this way, they may bundle their strengths and cover each other’s weaknesses.

Regarding the generalizability of this explorative study, several points should be noted. Token labeling is a specific task, and the behavior of direct and indirect GenAI annotators may differ for tasks of another nature. The prompts used for direct annotation have not been systematically tested, and it’s possible that especially zero-shot direct annotation would have better results with more guidance regarding the output format. The held-out test set was small and from the same document (i.e., not the same data) as the training data, which might have influenced the results. The training of the indirect annotation model has been done with just 500 examples, a typically low number for fine-tuning its underlying transformer model. Additionally, during the training of the indirect model, automatic reweighting was not applied (as we deemed its effects in the LLMaAA paper to be marginal), but integrating it might approve the model.

6. Conclusion

Despite these limitations, this study shows potential for applying GenAI as indirect annotators in humanities research. However, there are notable differences to other annotation strategies. Future research should address questions about indirect GenAI annotation's performance on other tasks (e.g., text classification), the impact of prompt optimizing frameworks (e.g., DSPy [13, 14]), and the potential of combining human and GenAI annotations to check on each other.

Author Contributions

Conceptualization: Arjan van Dalftsen, Folgert Karsdorp, Ayoub Bagheri, Els Stronks; Data Curation: Thirza van Engelen, Dieuwertje Mentink; Investigation: Arjan van Dalftsen; Methodology: Arjan van Dalftsen; Writing - Original Draft: Arjan van Dalftsen; Writing - Review & Editing: Folgert Karsdorp, Ayoub Bagheri, Els Stronks; Visualization: Arjan van Dalftsen; Supervision: Folgert Karsdorp, Ayoub Bagheri, Els Stronks.

Acknowledgments

This research would not have been possible without the financial support from Utrecht University AI Labs, the Meertens Instituut, and the Utrecht University focus area Advanced Data Science. Their generous contributions provided the necessary resources to conduct this study. Additionally, we would like to extend our gratitude to SURF for providing cloud computing services, which were instrumental in the analysis and processing of our data.

References

- [1] E. M. Arevalo and L. Fonteyn. "Non-Parametric Word Sense Disambiguation for Historical Languages". In: *Proceedings of the 2nd International Workshop on Natural Language Processing for Digital Humanities*. Taipei, Taiwan, 2022. URL: <https://aclanthology.org/2022.nlp4dh-1.16>.
- [2] J. Borst, J. Klähn, and M. Burghardt. "Death of the Dictionary?– The Rise of Zero-Shot Sentiment Classification". In: *Computational Humanities Research Conference (CHR)*. Paris, France, 2023, pp. 303–319. URL: <https://ceur-ws.org/Vol-3558/paper3130.pdf>.
- [3] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. *Language Models are Few-Shot Learners*. arXiv preprint <https://arxiv.org/abs/2005.14165>. 2020. DOI: 10.48550/arXiv.2005.14165.

- [4] Y. Chen, S. Li, Y. Li, and M. Atari. *Surveying the Dead Minds: Historical-Psychological Text Analysis with Contextualized Construct Representation (CCR) for Classical Chinese*. arXiv preprint <https://arxiv.org/abs/2403.00509>. 2024. DOI: <https://doi.org/10.48550/arXiv.2403.00509>.
- [5] T. Dejaeghere, P. Singh, E. Lefever, and J. Birkholz. “Exploring Aspect-Based Sentiment Analysis Methodologies for Literary-Historical Research Purposes”. In: *Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA) LREC-COLING-2024*. Torino, Italia: ELRA and ICCL, 2024. URL: <https://aclanthology.org/2024.lt4hala-1.16>.
- [6] Digitale Bibliotheek voor de Nederlandse Letteren (DBNL). *Collectie publiek domein*. <https://www.dbnl.org/letterkunde/pd/index.php>. 2023.
- [7] B. Ding, C. Qin, L. Liu, Y. K. Chia, B. Li, S. Joty, and L. Bing. “Is GPT-3 a Good Data Annotator?” In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Toronto, Canada, 2023. DOI: 10.18653/v1/2023.acl-long.626.
- [8] M. van Gompel. *python-ucto [computer software]*. <https://languagemachines.github.io/ucto/>. 2023.
- [9] R. Han, T. Peng, C. Yang, B. Wang, L. Liu, and X. Wan. *Is Information Extraction Solved by ChatGPT? An Analysis of Performance, Evaluation Criteria, Robustness and Errors*. arXiv preprint <https://arxiv.org/abs/2305.14450>. 2023. DOI: <https://doi.org/10.48550/arXiv.2305.01445>.
- [10] X. He, Z. Lin, Y. Gong, A.-L. Jin, H. Zhang, C. Lin, J. Jiao, S. M. Yiu, N. Duan, and W. Chen. *AnnoLLM: Making Large Language Models to Be Better Crowdsourced Annotators*. arXiv preprint <https://arxiv.org/abs/2303.16854>. 2024. DOI: <https://doi.org/10.48550/arXiv.2303.16854>.
- [11] A. Karjus. *Large language models to supercharge humanities and cultural analytics research*. Poster presentation at CHR2023 <https://2023.computational-humanities-research.org/programme/>. 2023.
- [12] A. Karjus. *Machine-assisted mixed methods: augmenting humanities and social sciences with artificial intelligence*. arXiv preprint <https://arxiv.org/abs/2309.14379>. 2023. DOI: <https://doi.org/10.48550/arXiv.2309.14379>.
- [13] O. Khattab, K. Santhanam, X. L. Li, D. Hall, percy Liang, C. Potts, and M. Zaharia. *Demonstrate-Search-Predict: Composing Retrieval and Language Models for Knowledge-Intensive NLP*. arXiv preprint <https://arxiv.org/abs/2212.14024>. 2022. DOI: <https://doi.org/10.48550/arXiv.2212.14024>.
- [14] O. Khattab, A. Singhvi, P. Maheshwari, Z. Zhang, K. Santhanam, S. Vardhamanan, S. Haq, A. Sharma, T. T. Joshi, H. Moazam, H. Miller, M. Zaharia, and C. Potts. *DSPy: Compiling Declarative Language Model Calls into Self-Improving Pipelines*. arXiv preprint <https://arxiv.org/abs/2310.03714>. 2023. DOI: <https://doi.org/10.48550/arXiv.2310.03714>.

- [15] J.-C. Klie, M. Bugert, B. Boullosa, R. E. de Castilho, and I. Gurevych. “The INCEPTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation”. In: *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*. Santa Fe, New Mexico, 2018. URL: <https://aclanthology.org/C18-2002>.
- [16] J. Kocoń, I. Cichecki, O. Kaszyca, M. Kochanek, D. Szydło, J. Baran, J. Bielaniewicz, M. Gruza, A. Janz, K. Kanclerz, A. Kocoń, B. Koptyra, W. Mieleszczenko-Kowszewicz, P. Miłkowski, M. Oleksy, M. Piasecki, Ł. Radliński, K. Wojtasik, S. Woźniak, and P. Kazienko. “ChatGPT: Jack of all trades, master of none”. In: *Information Fusion* 99 (2023), p. 101861. DOI: <https://doi.org/10.1016/j.inffus.2023.101861>.
- [17] Koninklijke Bibliotheek. *Over ons - Diensten DBNL*. <https://www.kb.nl/over-ons/dienst-en/dbnl>. 2024.
- [18] A. Kosar, G. D. Pauw, and W. Daelemans. “Comparative Evaluation of Topic Detection: Humans vs. LLMs”. In: *Computational Linguistics in the Netherlands Journal* 13 (2024), pp. 91–120. URL: <https://www.clinjournal.org/clinj/article/view/173>.
- [19] L. D. Langhe, A. Maladry, B. Vanroy, L. D. Bruyne, P. Singh, E. Lefever, and O. D. Clercq. “Benchmarking Zero-Shot Text Classification for Dutch”. In: *Computational Linguistics in the Netherlands Journal* 13 (2024), pp. 63–90. URL: <https://clinjournal.org/clinj/article/view/172>.
- [20] M. Li, T. Shi, C. Ziems, M.-Y. Kan, N. Chen, Z. Liu, and D. Yang. “CoAnnotating: Uncertainty-Guided Work Allocation between Human and Large Language Models for Data Annotation”. In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. Singapore, 2023. DOI: 10.18653/v1/2023.emnlp-main.92.
- [21] J. Liu, D. Shen, Y. Zhang, B. Dolan, L. Carin, and W. Chen. “What Makes Good In-Context Examples for GPT-3?” In: *Proceedings of Deep Learning Inside Out (DeeLIO 2022): The 3rd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures*. Dublin, Ireland and Online, 2022. DOI: 10.18653/v1/2022.deelio-1.10.
- [22] L. Molle. “Inleiding - Een geschiedenis van mensen en (andere) dieren”. In: *Tijdschrift voor Geschiedenis* 125 (2012), pp. 464–475. DOI: 10.5117/tvgesch2012.4.moll.
- [23] OpenAI. *Introducing ChatGPT*. <https://openai.com/blog/chatgpt>. 2022.
- [24] C. Qin, A. Zhang, Z. Zhang, J. Chen, M. Yasunaga, and D. Yang. *Is ChatGPT a General-Purpose Natural Language Processing Task Solver?* arXiv preprint <https://arxiv.org/abs/2302.06476>. 2023. DOI: <https://doi.org/10.48550/arXiv.2302.06476>.
- [25] S. Reborá, M. Lehmann, A. Heumann, W. Ding, and G. Lauer. “Comparing ChatGPT to Human Raters and Sentiment Analysis Tools for German Children’s Literature”. In: *Computational Humanities Research Conference (CHR)*. Paris, France, 2023. URL: <https://ceur-ws.org/Vol-3558/paper3340.pdf>.
- [26] N. Reimers and I. Gurevych. “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks”. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Hong Kong, 2019, pp. 3982–3992. DOI: <https://doi.org/10.48550/arXiv.1908.10084>.

- [27] M. Ren, W. Zeng, B. Yang, and R. Urtasun. *Learning to Reweight Examples for Robust Deep Learning*. arXiv preprint <https://arxiv.org/abs/1803.09050>. 2019. DOI: <https://doi.org/10.48550/arXiv.1803.09050>.
- [28] H. Ritvo. *The Animal Estate: The English and Other Creatures in the Victorian Age*. New ed. Cambridge, MA: Harvard University Press, 1989.
- [29] K. Thomas. *Man and the Natural World: Changing Attitudes in England 1500-1800*. New edition. London, UK: Penguin Books Ltd, 1991.
- [30] P. Törnberg. *ChatGPT-4 Outperforms Experts and Crowd Workers in Annotating Political Twitter Messages with Zero-Shot Learning*. arXiv preprint <https://arxiv.org/abs/2304.06588>. 2023. DOI: <https://doi.org/10.48550/arXiv.2304.06588>.
- [31] S. Wang, Y. Liu, Y. Xu, C. Zhu, and M. Zeng. “Want To Reduce Labeling Cost? GPT-3 Can Help”. In: *Findings of the Association for Computational Linguistics: EMNLP 2021*. Punta Cana, Dominican Republic, 2021. DOI: 10.18653/v1/2021.findings-emnlp.354.
- [32] R. Zhang, Y. Li, Y. Ma, M. Zhou, and L. Zou. “LLMaAA: Making Large Language Models as Active Annotators”. In: *Findings of the Association for Computational Linguistics: EMNLP 2023*. Singapore, 2023. DOI: 10.18653/v1/2023.findings-emnlp.872.
- [33] C. Ziems, W. Held, O. Shaikh, J. Chen, Z. Zhang, and D. Yang. *Can Large Language Models Transform Computational Social Science?* arXiv preprint <https://arxiv.org/abs/2305.03514>. 2024. DOI: <https://doi.org/10.48550/arXiv.2305.03514>.

7. Appendices

A. Annotation Guidelines

A.1. Annotation Schema

Category

- Animals: A living thing that can move around to search for food. It usually has ways to see, hear, smell, taste, and feel the world around it.
- Plants: A living thing that usually stays in one place. It creates its own food using sunlight, water, and air.

Type

- Organisms: A whole, living animal or plant. Think of it like one complete cat, or one whole oak tree.
- Parts: A piece of an animal or plant. Things like a bird’s wing, a flower petal, or a bear’s claw.
- Products: Something we get from a plant or animal that we use. Only first-order products count (i.e. it’s the first “product” that comes from the plant/animal, not a product of an earlier product). Examples are milk from a cow, honey from bees, or apples from a tree.

- **Collective:** Something is collective if the word refers to a heterogenous multitude of plants/animals. Nature explicitly and inherently is a prominent part of, but it is not 100% clear what kinds of nature. If the collective might belong to both categories (you choose the best or least-wrong category). Examples are: weide, grastapijt, bos, woud, vee, kudde.

Usage

- **Literal:** When the word means exactly the animal, plant, part, or product itself. If you envision the text, you should see it. ("The bear ate a fish.")
- **Symbolical:** When the word is used as a symbol or metaphor, representing something else. If you envision the text, you should not see it. ("His heart was as cold as a snake.") Pictures are symbolic. Nicknames are probably symbolical.
- **Petrified:** if the plants/animals word is the name of something or someone.

A.2. Technicalities

General rule Textual context is always dominant in annotating.

Discontinuous annotations Sometimes an annotation is discontinuous, meaning that there are words between the parts to be annotated. An example is: "esschen- en pijnhout". Here, "esschen-" and "hout" should be annotated. This can be done by annotating both (here, meaning that you should also make a separate annotation for hout!) and then defining a relationship. Step-by-step guide: 1. Annotate "esschen-" and "pijnhout" and "hout". 2. Select the first part ("esschen-"). 3. Right click the second part ("hout"). 4. Click "Link to" and select "discontinuous entity".

Part – Whole constructions Sometimes part-whole constructions occur, e.g. "de wortel van de brem". Here, it is important to look at the parts that are separately referential (wortel and brem, here). If the text has "bremwortel" there is just one separate referential entity.

Syntactic head Concerning compound words, we annotate based on the syntactic head. You can find the syntactic head by doing a reference test: to what part of the compound can you refer? ("hazenpad": not annotated; "padhaas": annotated). In Dutch, the syntactic head is normally on the right side of the word.

Co-references Co-references to entities are not tagged. (in "de wolf is blij, hij eet graag haas", "hij" should not be annotated). Likewise, words that in a specific context refer to plants/animals should not be annotated, unless the plant/animal aspect is inherent. "Veulen" and "kalf" are names for young animals and should be annotated, however, "jong", "wijfje", "mannetje", "wederhelft", "lichaam", are not.

Adjectives As a general rule, adjectives are not annotated. There are a few exceptions: 1. If the adjective is part of the name of a plant/animal, it should be annotated (e.g. "blauwe" in "blauwe vinvis" and "kruipende" in "kruipende boterbloem"). 2. Sometimes a word looks like an adjective, but it is used as a substantive. In that case, annotate it.

Foreign languages When plants/animals/nature-locations are in a non-Dutch language, they should still be annotated. There are two exceptions: 1. If the whole text is in a different language, it should not be annotated; 2. If the entities name is in a non-Latin script (e.g., Arabic, Greek, Hebrew), it shouldn't be annotated.

B. Used Prompts

A `ChatPromptTemplate` is used. Therefore, "chat messages" are provided between parentheses; inside the parentheses the "sender" of the message and the message itself are divided with a comma. Parts in *italics* dependent on the texts that is annotated. Here, it is only indicated that these parts exist.

B.1. Pre-filtering prompt

```
(
System,
You are a helpful assistant. You'll get a historical Dutch text.
It's your task to tell whether (non-human)
animals or plants are directly present in this text. You do this
by reasoning step by step, and then end by completing: 'I deem
the statement that literal animals are present in this text to
be:' with True or False. I know you can do it!
),
(
User,
\textit{Text to pre-filter}
)
```

B.2. Few-shot direct annotation prompt

```
(
System,

You are a highly intelligent and accurate nature domain information extraction
system. I'll provide a small text, written in historical Dutch. Your task is
to recognize and extract all entities related to plants or animals. If you
have found anything that falls into that category, you should annotate it on
three levels: 1. Category; 2. Type; 3. Usage.
```

For Category, there are two possibilities: Plants and Animals.

* Animals: A living thing that can move around to search for food. It usually has ways to see, hear, smell, taste, and feel the world around it.

* Plants: A living thing that usually stays in one place. It creates its own

food using sunlight, water, and air.

For Type, there are four possibilities: Organisms, Parts, Products, Collective.

- * Organisms: A whole, living animal or plant. Think of it like one complete cat, or one whole oak tree.
- * Parts: A piece of an animal or plant. Things like a bird's wing, a flower petal, or a bear's claw.
- * Products: Something we get from a plant or animal that we use. Only first-order products count (i.e. it's the first 'product' that comes from the plant/animal, not a product of an earlier product). Examples are milk from a cow, honey from bees, or apples from a tree.
- * Collective: Something is collective if the word refers to a heterogeneous multitude of plants/animals. Nature explicitly and inherently is a prominent part of, but it is not 100% clear what kinds of nature. If the collective might belong to both categories (you choose the best or least-wrong category). Examples are: weide, grastapijt, bos, woud, vee, kudde.

For Usage there are three possibilities: Literal, Symbolical, Petrified.

- * Literal: When the word means exactly the animal, plant, part, or product itself. If you envision the text, you should see it. ('The bear ate a fish.')
- * Symbolical: When the word is used as a symbol or metaphor, representing something else. If you envision the text, you should not see it. ('His heart was as cold as a snake.') Pictures are symbolic. Nicknames are probably symbolical.
- * Petrified: if the plants/animals word is the name of something or someone.

To summarize, you should detect all plant and animal related words and tag them according to this schema. So, for each found entity you annotate its category (Plant/Animal), its Type (Organisms/Parts/Products/Collective), and its Usage (Literal/Symbolical/Petrified).

It is extremely important that you work precise. Therefore, you should explain step by step why you make a choice. Also extremely important: the annotation you do should be in the form of a list with dictionaries. You should also do an explanation, but your ultimate annotation should be in that format. So you should always have output like this:


```
[{"span": span, "type": Category-Type-Usage}, ...]
```

Very important: if you don't find any entities, your annotation should be an empty dictionary in a list:

```
[{}]
```

otherwise the postprocess script will get in trouble.

Good luck, I count on you!

```
)  
(  
System,  
  The span must be exactly the same as in the original text, including white  
  spaces.  
)  
(  
User,  
Here are some examples:  
\textit{ Example1, Example2, Example3, Example4, Example 5}
```

Please now annotate the following input:

```
Input: \textit{Text to annotate.}  
)
```

B.3. Zero-shot direct annotation prompt

Same prompt as above, but without the examples.

C. Model Settings

OpenAI API parameters temperature = 1; top_p = 1; frequency_penalty = 0; presence_penalty = 0; gpt-4-o version: gpt-4o-2024-05-13. gpt-3.5-turbo version: gpt-3.5-turbo-0125. '2023-03-15-preview'

GysBERT parameters architecture: BertForTokenClassification; optimizer: Adam; learning_rate: 2e-5.

D. Held-out Test Set

1. Doch in kommerlyke tyden word dit kruid, een weinig geroost, door de menschen ten spyze gebruikt.
2. Al de gedroogde visch, die zich toen op het eiland bevond, werd daar van geheel zwart en onbruikbaar, ja in de twe naastvolgende jaren werden door die assche, of veeleer door

- de 'er mede vermengde scherpachtige rotsbrokjes of zand, gelyk boven by den brand op Jan Mayen eiland aangemerkt is, zo verre het
3. Als men het Varken, in 't midden aan weeder zyden van de rugge-graad, doorgesneden heeft, zo laat men ieder helft, even onder de schouder nog eens doorsnyden in de breedte.
 4. Het gerookt vleesch moet ook acht dagen in het zout leggen, en dan in zakken genaait in de rook gehangen worden, en moet drie of wel vier maanden rooken.
 5. Zouten van Spek, Hammen en Ossen-Vleesch, hoe daar mede te handelen.
 6. Dan legt men alles aldus in de kuip om in order te gebruiken: 1. de 6 klapstukken van de buik onder in, want ze kunnen het langste duuren: 2. de staartstukken: 3. de schouderbladeren: 4. de twee borststukken: 5. de twee beste ribben: 6. de vier andere ribben: de twee ongeschikte ribben die by de schouwers zitten: 7. de huspot zo men wil boven op; maar men moet zorg dragen dat de stukken wel vast in malkanderen sluiten, en de openingen moeten met zout gevuld worden, en wat zout 'er boven op, en eerst onder op den bodem gespreid; ook moet de kuip eerst schoon uitgebroid en met kruidnagels gedroogt worden.
 7. Dit alles te zaamen in een groote pan of styfse kom of hakkebord gedaan, en 6 tinne kommetjes met Osse-vleesch-nat of ander vleesch-nat, of warm water daar op gegooten en digt toegedekt en altemets eens omgeroert. en zo een nagt over, op de warme plaat laten staan weken; en dan stopt men ze gelyk Leverbeulingen; dog om dat de gort sterk zwelt maar half vol, en dan zynze half vol als men ze plat duuwt: Als ze gestopt zyn laat men ze zeer zagt kookken dat het water maar even beweegt omtrent een half uurtje, en men prikt ze ondertussen met een doortje om niet te barsten en uittekookken, en dan zyn ze heel goed. 4.
 8. En schoon veele staande houden dat het eeten van dit vleesch geen quaad aan de menschen doedt, zo zyn fatsoendelyke lieden nogtans beschroomd om het te gebruiken: om dit met zekerheid te weten zo kan men daar deeze proeve van neemen.
 9. §. XXXI. De Koemelk word tot artzeny gebruikt. De Melk is de voornaamste artzeny der Yslanders, en word daarom ook, zodra zy van de koe kooft, door gene anderen, dan alleen kranken, genoten.
 10. Die het beter willen maken, en 'er de middelen toe hebben, kopen een weinig zout, snyden, als het ge-slagt dier noch onafgehakt hangt, op drie of vier plaats en een diepe snede in het vleesch, en doen in iedere opening een kleine hand vol zout, zich verbeeldende, dat het dus zelf, zo veel nodig is, door het gantsche beest trekt, en het vleesch, wanneer 'er vervolgens wind en rook by kooft, zeer wel bewaard word Op de beide gezegde wyzen handelen de ingezetenen ook met het schapenvleesch, als zy het voor hun huisgezin slagten.
 11. Zeeusche Pens en Hoofdvliesch, hoe men die maaken zal.
 12. Ossen en Koeyen vallen niet groter dan het kleinst geestvee in Duitsland; hebben, gelyk bereids gezegt is, gene Hoornen, en genieten alleen het voorrecht, door de huis lieden in den winter mede onder 't dak genomen en met het zo kommerlyk gewonnen hooy, of, by mangel van het zelve, met het gedroogd zeegewas Zeenestel spaarzaam gevoed te worden.

13. Men stopt de beulingen maar half vol om dat die anders te ligt uitkooken of barsten; en men bind ze met een touwtje onder en boven toe, en dan wordenze op een schootel plat nedergelegd, tot dat ze gekookt worden: Voor al moet men niet vergeeten genoeg vet daar in te doen, want anders zyn de Leverbeulingen te droog.
14. weshalven de boeren 'er aldaar meer acht op geven. Dezen jagen alleen de Hamels in 't gebergte; doch houden de Oyen zo veel by huis, als doenlyk is.
15. Als men zo veel moeiten niet doen wil om Rolpens en Hoofd-vleesch te maaken, zo snyd men de pens in stukken, en men kookt het met de kop tot dat alles gaar is, en dan legt men het vleesch met de pens door een, met wat zout en heele peper, in den azyn, in een keulse aarde pot, is heel goed om met appelen des winters gebakken te eeten. 17.
16. De Boter kaarnen de meesten voor en na zo hairig, als zy uit ongereinigde melk in een zamengenaaide schapenvacht gemolken is, en leggen dezelve dus op; weshalven een vreemdeling die Boter niet ligtelyk door de keel zoude kunnen krygen.
17. Dan neemt men een groote vleesch keetel en men hangt ze vol regen water over het vuur, en als het water kookt doet men de beulingen daar in, dat die regt uit en niet op malkanderen leggen, daarom mag men niet meer als anderhalf douzyn beulingen te gelyk kooken; en ze moeten heel zagtjes kooken, omtrent een half uur lang.
18. Afhakken van 't vleesch in de Slacht-tyd, en hoe men de stukken best en ten meesten voordeelen zal gebruiken, en hoe men verder met alles in de Slacht-tyd, moet handelen. 1.
19. Neemt by de 20 ponden, gehakt redelyk vet, varkens vleesch, anderhalf loot of twee loot nootemuscaten; twee loot nagelen; twee loot zwarte peeper, dit alles ter deegen fyn gestooten zynde, zo roert men het onder anderhalf vierendeel zout, en men kneed het door het gekapte Varkensvleesch heen; en men laat het zo een nacht met een schoone doek bedekt staan doortrekken.
20. Hunne vellen vallen in den winter, als zy het meeste en vastste hair hebben, het best; weshalven de Yslanders dezelve dan naarstig vangen, en wel, uit aangebore afschuw van schietgeweer, met uitgezette netten of vangyzers, die gelyk een kleermakersschaar gevormt, en met een dood lam ten lokaas voorzien zyn.
21. geweld der uitbrekende en uitgezette lucht een groot gedeelte van den berg, 't geen te zwaar was, om opgeligt te worden, op zyde en niet slegts een gantsche myl wegs langs het eiland tot aan het strand, maar zelfs noch een myl verr' in zee voortgeschoven, en aldaar neder gezet wierd, alwaar het, onaangezien de diepte, in den beginne wel 60 vademmen boven het water uitstak, en aldaar merendeels noch staat e.
22. Neemt voor het vleesch, het geen men daar in legt, het vleesch van de schouwer van een Os dat het malste is; of anders een van de platte billen.
23. Ja zy zyn het zelven, die gemeenlyk het begin der aardbranden veroorzaken.
24. §. XXXIV. Hebben geen Zwynen, maar wel Honden en Katten.
25. Doch wat de eigentlyke en natuurlyke oorzaak dezer zeldzaamheid zyn mag, is niet zeer ligt te beseffen w.
26. Reusel, hoe men die wel zal smelten.
27. Van harde of Coraalachtige Zeegewassen wist myn berichter te zeggen, dat enigen van dezelve op de gronden gevonden wierden; doch konde hen niet noemen of beschryven, nadien hy, volgens zyne eigen belydenis, 'er nooit naar gezien had.

28. Dezen zyn de Snoriper op de lappische Alpen, die zich a steeds op het land houden, meer lopen dan vliegen, en mitsdien niet bezwaarlyk te vangen zyn.
29. Men moet zich verwonderen, wat zy kunnen uitstaan; doch zy worden wel degelyk door de ongemakken verhard, nadien zy jaar uit jaar in in het open veld onder den bloten Hemel blyven, en 's winters onder de sneeuw zowel, als 's zomers, hun voeder zelve moeten zoeken, waar toe zy alleen de weldaad van de natuur genieten, dat zy met byzondere styve, lange en dikke hairen, allermeeft tegen den wintertyd, bedekt zyn.
30. Vervolgens bragt men het zieke volk aan land, 't geen, ofschoon het, behalven enig Lepelblad, niet als Zuring in warme Melk en een weinig Schapenvleesch nuttigde, nochtans velen binnen acht en de anderen binnen veertien dagen zo fris en gezond werden, dat zy huppelden en sprongen, en in minder dan vier weken na hun komst weder scheep gaan, zelve hun anker lichten, en die lange en bezwaarlyke reize voorts vrolyk voleinden konden.
31. Het vleesch snyd men eerst aan stukken als Ossekarbenaden; en dan snyd men het aan lange reepen omtrent een vinger dik en vierkant; men snyd het vet ook aan zulke langwerpige stukken; en dan bestrooid men de pens met wat geprepareerd zout en kruid, gelyk ik boven gezegt heb.
32. Men neemt 3 loot bruine peper, en een halfvierendeel nagelen; dit te zaamen eerst fyn gestoten en in een aarde schootel gedaan, en een hand vol gedroogde Saly, die men op den haart wat te droogen legt en die klein gewreven is, en een hand vol of vier zout daar onder geroert, tot men denkt dat men genoeg zal hebben; want den een doet het wel wat hartiger dan den ander.
33. Men behoeft 'er geen Sukade nog Amandelen in te doen als men niet wil, en is evenwel goed maar zo lekker niet.
34. De Harsten laat men een dag of vyf in het zout leggen en men moet ze niet te groot laten hakken, om dat ze anders te ongeschikt zyn, en ieder een doet dit naa de groote van zyn huisgezin, ook zyn die Harsten dus zeer goed om in den Oven gezet en gebraden te worden.
35. Het vleesch om in de Kuip in te zouten, daar neemt men toe de zes klapstukken van de buyk, de twee staartstukken, de schouwderbladeren, de twee borststukken, de vier andere ribben, als men de twee beste ribben wil in de rook hangen, anders kan men ook de Paterstukken inzouten, en dan nog de twee ongeschikte ribben die by de schouders zitten; en men laat die stukken groot of klein hakken naa dat men het wil hebben en het huisgezin groot is.
36. Mitsdien ziet men zelden op Ysland andere, dan uitgebrande bergen, aan en om welke men bequaam de werkingen en overgebleven tekenen van een vorigen brand bespeuren kan.
37. Buiten dien tyd leggen de inwoonders, nadien de Vossen de schapen zeer schadelyk zyn, kraanogen (*nucis vomicae*) in honig geweekt, die zy, anders niets zoets te eten bekomende, zeer begerig inzwelgen.
38. Neemt 4 kop Gort schoon afgewasschen: 4 pond korenten die wel verlezen en schoon gewassen zyn: 8 loot gestoote kaneel: 1 loot gestoote nagelen: 3 loot gestoote notemuscaaten; 1/2 pond poeijer-zuiker: 1 pond gepelde amandelen in stukjes gesneden: 6 sukade

schellen aan stukjes gesneden: Een hand vol zout: 10 pond of daar omtrent Osse-niervet aan dobbelsteentjes gesneden.

39. Het zoude gezwollen verwekken, en, als men 'er veel van eet, sterk openende zyn.
40. de Ravens verjagen; doch het Lam, vermits het, zyn voeder niet konnende zoeken, elendig omkomen moet, slagten, en het het zachte vel afstropen, 't geen de peltery geeft, die in Denmarken en Holstein onder den naam van Schmaaskin of Schmaasken x verkogt en zeer veel door lieden van een middelbaar vermogen gedragen word.
41. Neemt een van de grootste Kalfskoppen, en reinigt die, en wascht ze vier of vyfmalen ter degen schoon af, en laatze een nacht in schoon regen water staan te trekken, dat 'er de slym en het bloedige wel schoon af is, en hangt de kop met schoon regen-water over het vuur; en doet ook in de keetel, de nek van het Varken, en de twee ooren met wat veel zwoort dat 'er genoeg is om het vleesch in het hakkebord van boven en onderen te bedekken; en als men te veel zwoort en niet genoeg vleesch heeft, zo doet men 'er wel een of twee van de vleesigste stukken van het varken by, en men laat het te zaamen een uur of drie kooken, na dat men het alvorens wel schoon geschuimt heeft, en het moet zeer gaar zyn tot dat het vleesch van de beenen af valt, en dan schept men het uit op een aarde vergiettest of doorslag.
42. Hunne manier, om het Rundvee te slagten, heeft ook iets byzonders, Zy kollen het niet voor den kop, menende, dat daar door het bloed in 't vleesch stremt, en mitsdien niet lopen kan; maar steken het een dun penmes diep in den nek, waar door het ter aarde valt; als dan trekken zy de poten gezwind met strikken zamen, en openen de keel, op dat al het bloed zoude uitvlieten Het ingewand word door de Yslanders allereerst, zonder veel te reinigen, genuttigt, en het dier zelf afgehakt.
43. Neemt de Lever van het Varken en wascht die schoon, en laat die op een aarde schotel leggen; doet daar zo raauw de vellen en spieren met een mes ter degen schoon uit, en doet het in een schoon tobbetje.
44. Voor een geheele pens heeft men omtrent 20 pond vleesch noodig, behalven het vet dat men daar by gebruikt, dat nog omtrent 10 ponden is.
45. Laat dan een ketel of twee met regenwater kooken en laat het Koud worden; en als het koud is neemt dan schaars drie kommetjes van dat water tegen ruim een kommetje wyn azyn, en mengt dat te zaamen onder malkanderen zo veel tot dat de pens, als he daar over gegooten is, kan onderleggen, en zet ze dan zo open weg daar ze niet te vogtig staan, is heel goed om met appelen gebakken, of gestooft met wyn te eeten. 10.
46. Saucysen of Worst van Varkenvleesch, hoe men die maaken zal.
47. De stukken worden niet met zout gewreven, maar slegts twemaal door zeewater gehaalt, en dan in de lucht, op dat zy winddroog zouden worden, en vervolgens in hunne hutten over hunne haardsteden gehangen, om dezelve te roken, en te meer te doen drogen Dus behandelen zy hun geslagt half verrot en half stinkend vleesch, tot zy het voorts opeten.
48. Het vleesch om in de rook te hangen daar toe neemt men de Paterstukken, de twee andere platte billen; en de twee beste ribben; en de spieren, die achter tusschen de beenen van de ribben inzitten, moeten daar schoon uitgedaan worden, om dat daar door ligt verderf ontstaan kan.

49. Dan doet men twee geraspte nootemuscaten, met wat gestoote foelie en met wat zout daar in, en men hakt het te zaamen onder een tot het redelyk klein, maar niet al te klein is.
50. Alsdan begeeft een harder zich met de afgerichte honden op een heuvel, en geeft met zyn hoorn een teken, waarop de honden zich verdelen, en de Schapen van alle kanten uit de klippen en wildernissen in een zekere omtuining of staketzel dryven, 't geen vooraan wyd uitgezet is; doch, op dat zy niet zouden kunnen ontvluchten, naar achter allengs enger word.

E. Cost

In humanities, costs are often an important consideration. For all strategies, there's the cost of establishing annotation guidelines and making something of a test set. After that point:

- Human annotation costs €0.7 per sentence;
- Direct annotation costs €0.007 per sentence (for GPT-4o, directly via OpenAI), with zero-shotting being slightly cheaper due to omitting examples in the prompt;
- Indirect annotation costs €4.50 to train the model, and nothing per sentence.

It is important to emphasize that costs per strategy are likely to change over time since GenAI models are getting cheaper and that human annotations might differ significantly per country or institution. Also, multiple factors should be considered, such as available hardware and environmental effects.

F. Online Resources

Code and data used in this study can be found here:

- Data: <https://www.dbnl.org/letterkunde/pd/index.php>,
- Code and annotations GitHub Repository.