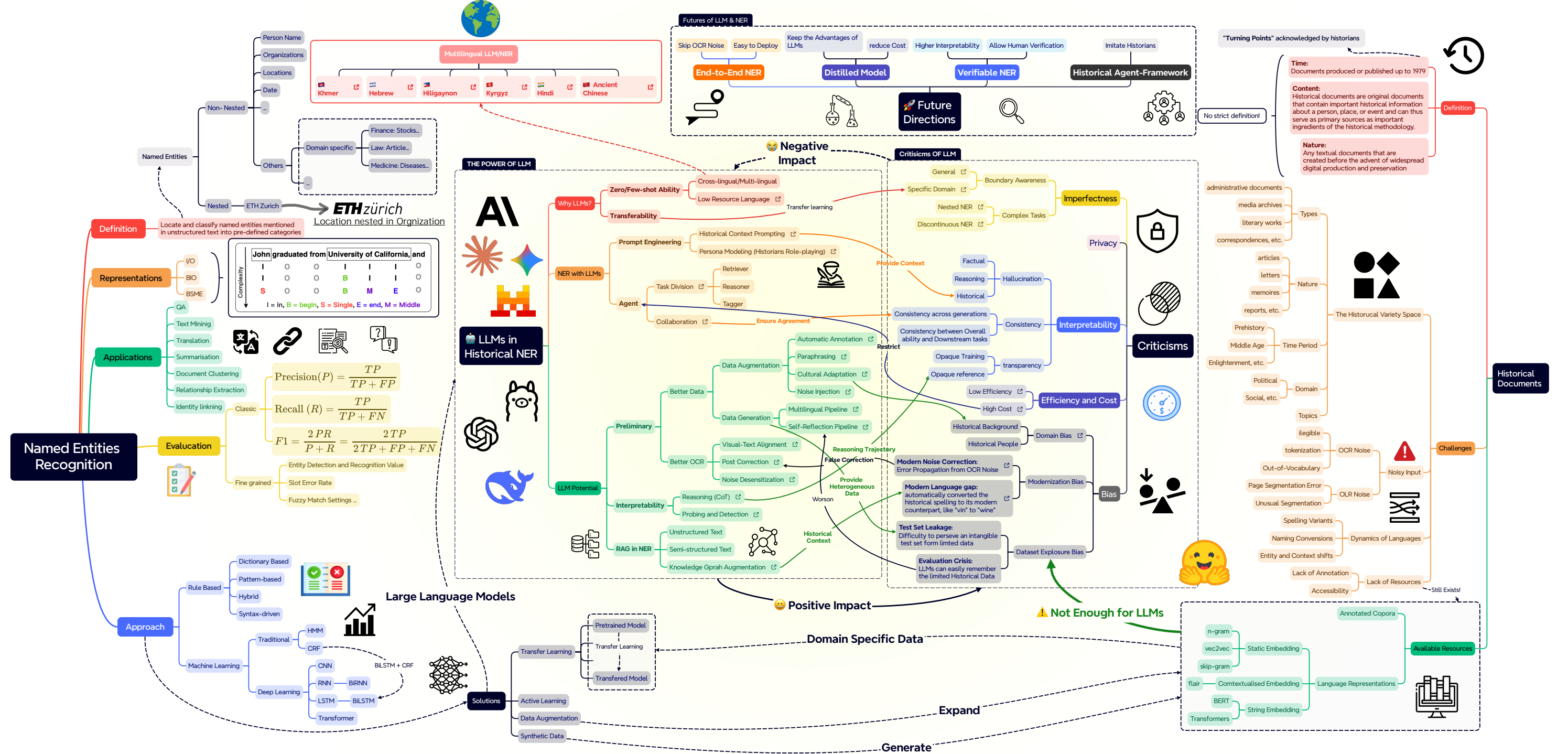


Named Entity Recognition and Linking in Historical Texts

In the Era of Large Language Models

Made by Shen Jiajun



Bibliography

[1] Ehrmann, M., Hamdi, A., Pontes, E. L., Romanello, M., & Doucet, A. (2023). Named entity recognition and classification in historical documents: A survey. *ACM Computing Surveys*, 56(2), 1-47.

[2] Zhang, S., & Colavizza, G. (2025). Named Entity Recognition of Historical Text via Large Language Model. *arXiv preprint arXiv:2508.18090*.

[3] Tudor, C., Megyesi, B., & Ostling, R. (2025, May). Prompting the Past: Exploring Zero-Shot Learning for Named Entity Recognition in Historical Texts Using Prompt-Answering LLMs. In *Proceedings of the 9th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLIL 2025)* (pp. 216-226).

[4] Hillmann, T., Dröge, M., Dresselhaus, N., Grallert, T., Althage, M., Bayer, P., ... & Skibba, A. (2025). NER4all or Context is All You Need: Using LLMs for low-effort, high-performance NER on historical texts. *A humanities informed approach*. *arXiv preprint arXiv:2502.04351*.

[5] Wang, Z., Chen, H., Xu, G., & Ren, M. (2025). A novel large-language-model-driven framework for named entity recognition. *Information Processing & Management*, "62"(3), 104054.

[6] Tao, X., Dong, X., & Zhou, X. (2025). OEIMA: Ontology-Enhanced Multi-Agent Collaboration Framework for Zero-Shot Clinical Named Entity Recognition. *arXiv preprint arXiv:2511.15211*.

[7] Naraki, Y., Yamaki, R., Ikeda, Y., Horie, T., Yoshida, K., Shimizu, R., & Naganuma, H. (2024). Augmenting NER datasets with LLMs: towards automated and refined annotation. *arXiv preprint arXiv:2404.01334*.

[8] Rengarajan, N. K., Yan, J., & Wang, C. (2025). PANER: A Paraphrase-Augmented Framework for Low-Resource Named Entity Recognition. *arXiv preprint arXiv:2510.17720*.

[9] Ehsan, T., & Solorio, T. (2025). Enhancing NER Performance in Low-Resource Pakistani Languages using Cross-Lingual Data Augmentation. *arXiv preprint arXiv:2504.08792*.

[10] Ye, J., Xu, N., Wang, Y., Zhou, J., Zhang, Q., Gui, T., & Huang, X. (2024). Llm-da: Data augmentation via large language models for few-shot named entity recognition. *arXiv preprint arXiv:2402.14568*.

[11] Gladstone, C., Fang, Z., & Stewart, S. D. (2025). Ground Truth Generation for Multilingual Historical NLP using LLMs. *arXiv preprint arXiv:2511.14688*.

[12] Heng, Y., Deng, C., Li, Y., Yu, Y., Li, Y., Zhang, R., & Zhang, C. (2024). Progen: Generating named entity recognition datasets step-by-step with self-reflexive large language models. *arXiv preprint arXiv:2403.11103*.

[13] Wei, H., Sun, Y., & Li, Y. (2025). Deepseek-ocr: Contexts optical compression. *arXiv preprint arXiv:2501.18234*.

[14] Boros, E., & Ehrmann, M. (2024, December). Investigating OCR-Sensitive Neurons to Improve Entity Recognition in Historical Documents. In *International Conference on Asian Digital Libraries* (pp. 54-66). Singapore: Springer Nature Singapore.

[15] Kanerva, J., Ledins, C., Kopyah, S., & Ginter, F. (2025). Ocr error post-correction with llms in historical documents: No free lunches. *arXiv preprint arXiv:2502.01205*.

[16] Huang, H., Chen, H., Huang, R., Lin, C., & Qin, Y. (2025). A Reasoning Paradigm for Named Entity Recognition. *arXiv preprint arXiv:2511.11978*.

[17] Morand, V., Tomeh, N., Mothe, J., & Piwowarski, B. (2025). ToMMeR--Efficient Entity Mention Detection from Large Language Models. *arXiv preprint arXiv:2510.19410*.

[18] Freund, M., Dorsch, R., Schmid, S., Wehr, T., & Harth, A. (2024, December). Enriching RDF data with LLM based named entity recognition and linking on embedded natural language annotations. In *International Knowledge Graph and Semantic Web Conference* (pp. 109-122). Cham: Springer Nature Switzerland.

[19] Kong, M., Buoy, R., Chenda, S., & Taing, N. (2025). Khmer Spellchecking: A Holistic Approach. *arXiv preprint arXiv:2511.09812*.

[20] Scheible-Schmitt, R. (2025). HalleluBERT: Let every token that has meaning bear its weight. *arXiv preprint arXiv:2510.21372*.

[21] Ald Teves, J., Cal, R. D., Magdiel Villaluz, J., Malolos, J., Magtira, M., Rodriguez, R., ... & Marvin Imperial, J. (2025). HiligayNER: A Baseline Named Entity Recognition Model for Hiligaynon. *arXiv e-prints*, arXiv:2510.

[22] Turatali, T., Alekseev, A., Jumalieva, G., Kabaeva, G., & Nikolenko, S. (2025, September). Human-Annotated NER Dataset for the Kyrgyz Language. In *2025 10th International Conference on Computer Science and Engineering (UBMK)* (pp. 1607-1612). IEEE.

[23] Singh, S., Mishra, R., & Tiwary, U. S. (2025). Enhancing Hindi NER in Low Context: A Comparative study of Transformer-based models with vs. without Retrieval Augmentation. *arXiv preprint arXiv:2507.16002*.

[24] Tang, X., Deng, Z., Su, Q., Yang, H., & Wang, J. (2024). CHIEC: an information extraction corpus for ancient Chinese history. *arXiv preprint arXiv:2403.15088*.

[25] Guo, Q., Dong, Y., Tian, L., Kang, Z., Zhang, Y., & Wang, S. (2025, January). BANER: Boundary-aware LLMs for few-shot named entity recognition. In *Proceedings of the 31st International Conference on Computational Linguistics* (pp. 10375-10389).

[26] Lu, Q., Li, R., Wen, A., Wang, J., Wang, L., & Liu, H. (2025, May). Large language models struggle in token-level clinical named entity recognition. In *AMIA Annual Symposium Proceedings* (Vol. 2024, p. 748).

[27] Kim, H., Kim, J. E., & Kim, H. (2024, November). Exploring nested named entity recognition with large language models: Methods, challenges, and insights. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 8653-8670).

[28] Yang, Y., Ma, F., Meng, S., Liu, A., & Wen, L. (2025). GapDNER: A Gap-Aware Grid Tagging Model for Discontinuous Named Entity Recognition. *arXiv preprint arXiv:2510.10927*.

[29] Lu, J., Wang, Y., Yang, Z., Liu, X., Mac Namee, B., & Huang, C. (2024). PaDeLLM-NER: parallel decoding in large language models for named entity recognition. *Advances in Neural Information Processing Systems*, "37", 117853-117880.

[30] Zhang, Z., Zhao, Y., Gao, H., & Hu, M. (2024, May). Linker: Linking local named entity recognition models to large language models using uncertainty. In *Proceedings of the ACM Web Conference 2024* (pp. 4047-4058).

[31] Piryani, B., Mozafari, J., Abdallah, A., Doucet, A., & Jatowt, A. (2025, November). Evaluating Robustness of LLMs in Question Answering on Multilingual Noisy OCR Data. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management* (pp. 2366-2376).