



**HAL**  
open science

# The Growing Importance of Humanities for NLP in the Era of LLMs

Mika Hämäläinen, Emily Öhman, So Miyagawa, Khalid Alnajjar, Yuri Bizzoni, Jack Rueter, Niko Partanen

## ► To cite this version:

Mika Hämäläinen, Emily Öhman, So Miyagawa, Khalid Alnajjar, Yuri Bizzoni, et al.. The Growing Importance of Humanities for NLP in the Era of LLMs. Lightning Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities, Nov 2024, Miami, United States. pp.2-6. <hal-04763819>

**HAL Id: hal-04763819**

**<https://hal.science/hal-04763819v1>**

Submitted on 2 Nov 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# The Growing Importance of Humanities for NLP in the Era of LLMs

*Mika Hämäläinen<sup>1</sup>, Emily Öbman<sup>2</sup>, So Miyagawa<sup>3</sup>, Khalid Alnajjar<sup>4</sup>, Yuri Bizsoni<sup>5</sup>, Jack Rueter<sup>6</sup> and Niko Partanen<sup>6</sup>*

*<sup>1</sup>Metropolia University of Applied Sciences*

*<sup>2</sup>Waseda University*

*<sup>3</sup>University of Tsukuba*

*<sup>4</sup>F-Secure Oyj*

*<sup>5</sup>Aarhus University*

*<sup>6</sup>University of Helsinki*

---

## Introduction

After the introduction of large language models (LLMs), science has not remained the same. Researchers from several different fields of science have been rushing to conduct research on LLMs. This is due to the fact that LLMs are no longer something only machine learning experts can understand. As the middle *L* in *LLM* stands for *language*, it is evident that humanities with their long tradition in studying language plays an increasingly important role in studying LLMs.

NLP4DH<sup>1</sup> started as an event that had one clear goal - to bring digital humanities (DH) and natural language processing (NLP) researchers together. Back in 2021, we perceived a gap between DH and NLP (see Hämäläinen et al., 2021).

---

<sup>1</sup> <https://www.nlp4dh.com>

DH researchers were more interested in the phenomenon they were studying than the actual computational methods. As a result, their methods were often quite traditional. NLP researchers were more interested in pushing the boundaries of the state of the art in machine learning. With the emergence of LLMs, this situation has changed a lot.

LLMs represent the state of the art in NLP, and due to the availability of easy to use LLMs and their APIs such as ChatGPT, Claude and Gemini, even people with no expertise in NLP or machine learning can easily embrace them in their research. This has introduced a paradigm shift in both DH and NLP research. As we can witness, this year NLP4DH has quite a few papers where LLMs are used for purely DH motivated research (see Hämäläinen et al., 2024).

While NLP researchers panicked at first when LLMs were introduced - something that was described as an existential crisis (see Saphra et al., 2023), DH researchers got a new set of powerful tools in their arsenal. This also entails that the gap between NLP and DH got much smaller, and yet interdisciplinary venues like NLP4DH are now more important than ever.

---

## Why Do Humanities Matter?

We have inarguably entered a new era in AI research. As tech leaders are taking turns in flashing the potential of artificial general intelligence (AGI) becoming a real thing sooner than we may think<sup>2,3,4</sup>, it is becoming more and more evident that how we conduct AI research has to change as well.

---

<sup>2</sup> <https://www.reuters.com/technology/teslas-musk-predicts-ai-will-be-smarter-than-smartest-human-next-year-2024-04-08/>

<sup>3</sup> <https://www.cnn.com/2024/01/16/openai-sam-altman-agi-coming-but-is-less-impactful-than-we-think.html>

<sup>4</sup> <https://fortune.com/2023/05/03/google-deepmind-ceo-agi-artificial-intelligence/>

We can no longer embrace the current NLP way of doing research. We cannot simply just measure scientific advancements by looking at numerical improvements on some benchmarking metrics. Not only are benchmarks unreliable (see Reiter, 2018; Bowman & Dahl, 2021; Van der Wal et al., 2024), intelligence in LLMs is emergent (see Chowdhery et al., 2023). LLMs possess a variety of different skills - they are jacks of all trades so to say. The fact that nobody can really give a definitive list of all the things a particular LLM can do also means that we cannot have a benchmark that could fairly rank LLMs.

We need to approach this new generation of AI from a completely different perspective. This is the perspective of humanities. We can no longer only measure LLMs quantitatively, but we also need qualitative and theoretical understanding. We all know that ChatGPT can translate between most big languages, and it does so remarkably well. The question is no longer whether ChatGPT o1 has a higher BLEU score than ChatGPT 4o. Instead, we are dealing with intelligent systems that should rather be studied from the point of view of translation studies.

Anything LLMs are and can do, where they excel and what mistakes they make are better understood from the point of view of the humanities than benchmarks. This is why bringing humanities and machine learning - DH and NLP - closer together is more important now than ever.

---

## **To Conclude**

Natural Language Processing for Digital Humanities (NLP4DH) stands at the forefront of a revolutionary intersection between technology and the human experience. As we peer into the future, the possibilities are thrilling and limitless.

The thirst for innovation in this field is palpable. Traditional NLP evaluation methods have been pushed aside by the rapid rise of LLMs, leaving a critical gap in our understanding. This is where the humanities come into the spotlight alongside the social and behavioral sciences.

With their deep insights into human nature, culture, and behavior, these disciplines are now indispensable. These disciplines provide the nuanced perspectives needed to understand and shape the behavior of LLM and future AI systems. nlp4dh bridges the gap between cold algorithms and warm human experience, allowing technology to process language and the rich expression of humans. It promises to usher in an era of genuinely understanding the tapestry.

As we move into this uncharted territory, the fusion of computational power and human insight will undoubtedly lead to breakthroughs unimaginable today, and the future of NLP4DH is not only bright, it is full of possibilities. We are excited to see where NLP4DH will take us in the future. The demand is there. Current evaluation methods in NLP are not sufficient in the era of LLMs. Humanities and closely related fields such as social sciences and behavioral sciences are now needed to better understand how LLMs and future AI behave.

---

## References

Bowman, S., & Dahl, G. (2021). What Will it Take to Fix Benchmarking in Natural Language Understanding?. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4843-4855).

Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., ... & Fiedel, N. (2023). Palm: Scaling language

modeling with pathways. *Journal of Machine Learning Research*, 24(240), 1-113.

Hämäläinen, M., Alnajjar, K., Partanen, N., & Rueter, J. (2021). Proceedings of the Workshop on Natural Language Processing for Digital Humanities. Association for Computational Linguistics.

Hämäläinen, M., Öhman, E., Miyagawa, S., Alnajjar, K., Bizzoni, Y. (2024). Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities. Association for Computational Linguistics.

Reiter, E. (2018). A structured review of the validity of BLEU. *Computational Linguistics*, 44(3), 393-401.

Saphra, N., Fleisig, E., Cho, K., & Lopez, A. (2023). First tragedy, then parse: History repeats itself in the new era of large language models. *arXiv preprint arXiv:2311.05020*.

Van der Wal, O., Bachmann, D., Leidinger, A., van Maanen, L., Zuidema, W., & Schulz, K. (2024). Undesirable biases in NLP: Addressing challenges of measurement. *Journal of Artificial Intelligence Research*, 79, 1-40.