# Novel or Drivel? Variants of Invariants for Teaching NLP in the LLM Era

**Marius Micluța-Câmpeanu**
Interdisciplinary School of Doctoral Studies, University of Bucharest, Romania
marius.micluta-campeanu@unibuc.ro

## Abstract

The ubiquitous adoption of large language models by students prompts teachers to redesign courses and evaluation methods, especially in computer science and natural language processing (NLP) where the impact is more tangible.

Our contribution is two-fold. First, we attempt to define invariants for the role of education itself given the over-abundance of information that appears to be more accessible than ever before. Then, we present our approach and materials used for an introductory course in NLP for undergraduate students, drawing inspiration from software engineering best practices. Our vision regarding large language models is to rely on local models to cultivate a sense of ownership and sovereignty in an age where every bit of independence and privacy get eroded.

## 1 Introduction

With the release of large language models (LLMs) to the general public in late 2022, there has been an ever-growing pressure on the education sector to fundamentally rework what is being taught and how students should be evaluated. Unfortunately, three years later, at least in our experience, initiatives in this direction comprise of minor course adjustments or ineffective efforts to thwart cheating. We commend the efforts of teachers that are adapting to this challenging environment, noting that it is difficult, if not impossible, to quantify the extent of these actions at scale.

Based on informal discussions with pupils and students of varying ages, we observed that younger generations understand the potential of these tools and frequently use (and misuse) them, in line with data reported in the literature (Naganuma et al., 2025; Kann, 2025), in spite of official interdiction. If we corroborate this with the fact that LLMs (both commercial and open-weight) are often capable of solving many routine programming tasks, continuing to impose restrictions on students is in our opinion both ineffective and counter-productive. We believe that teaching should be collaborative, not adversarial, and if students repeatedly encounter perceived superiority of LLMs, they will begin to question the credibility of their instructors. Through the lens of software engineering, current approaches could be seen as legacy systems that must be gradually replaced, since a complete rewrite would lead to major disruptions. Likewise, AI literacy should also be introduced gradually (Zhong and Liu, 2025).

## 2 Teaching Invariants

Playing devil's advocate, we argue that formal studies in computer science and NLP are no longer relevant, as every possible question one might have can be readily answered by a language model[1] at a lower cost,[2] in a shorter time span, and tailored to the learner's needs. In this light, we must start from invariants of the role of education that hold true throughout history by attempting to answer questions adjusted from the philosophy of education (Curren, 2025): "What do we want to achieve from education?", "What are our responsibilities regarding NLP usage and development?", and "How should we reach these goals?". These observations are central in our field due to the links between machine learning, language, and semantics, upon which students should ponder. As software continues to grow in complexity, writing all code by hand without any kind of automated aid (e.g. generating boilerplate, code review) will be less frequent for common tasks, allowing us to focus on higher-level issues.

While the above questions lend themselves to subjective answers, we believe they represent a key

[1]Even when the answers are wrong, they can still be helpful. We make these assumptions and claims as a worst-case hypothetical scenario.

[2]Smaller models can run on consumer hardware at decent speed even under 100W without requiring a discrete GPU.



*Proceedings of the Seventh Workshop on Teaching Natural Language Processing (TeachNLP 2026)*, pages 129–133
March 29, 2026 

starting point of what can and should be taught, since low-level aspects will be easier to automate. Specifically in the context of LLMs, our position is that these systems need to be regarded as tools, with humans *always* in the loop, reviewing and taking the final decision. Like many other tools, LLMs are amenable to dual-use, but our ideal must be to treat the root cause, not to blame the tools. Therefore, we should strive to instill moral values, encourage professionalism and responsibility, develop critical thinking, and stress the importance of verifying information from primary sources.

Previous editions of the TeachingNLP Workshop resulted in high quality resources that cover most technical aspects (Nikishina et al., 2024; Aßenmacher et al., 2024), though there is little work on the foundational aspects mentioned above, specifically when teaching NLP (Friedrich and Zesch, 2021; McDermott et al., 2025).

## 3 Proposed Curriculum

The materials discussed in this paper are for an optional introductory course to NLP for computer science undergraduate students in their last semester, spanning 10 weeks. The students are familiar with machine learning algorithms, neural networks, calculus, linear algebra, and statistical methods from prior courses.

The theoretical foundation is provided by Jurafsky and Martin (2025). The covered concepts, outlined below, unveil classical NLP techniques in a chronological order as an explanation of the reasons that led to deep learning and large language models, slightly similar to Joshi et al. (2024). Throughout the semester, we intersperse each concept in the curriculum with recent papers from our group presented as case studies, where these traditional methods found their use. We want to challenge the narrative that meaningful research in this field is impossible without high computational costs. Our hope is to orient students towards research by presenting current work that combines classical NLP with the ever-present LLMs. This way, we also intend to answer the question of why LLMs are in many situations not the right tool for the job.

Since our faculty has a graduate studies program specifically focused on NLP and since we only have usually only 8 or 9 weeks (due to national holidays), many other topics are inevitably left out from this course. For instance, training a large language model from scratch is studied as part of a course in the master's program.

### 3.1 Covered Concepts

We start by creating a dataset from scratch in order to show every part of the process. We intentionally choose a language other than English to raise awareness of issues with lower-resource languages.

This is followed by exploratory data analysis, data cleaning, and other forms of normalization, such as lemmatization and stemming. Next, to be able to use a learning algorithm to learn from data, we need to represent our inputs as features. We present sparse encodings and its drawbacks, serving as a motivation for dense representations. The story repeats with static embeddings, paving the way for contextual embeddings and transformers.

Last but not least, computational aspects of morphology are discussed. These include part-of-speech tagging (POS), dependency parsing, and named entity recognition (NER).

Time permitting, supplementary materials focus on aspects related to transformers, large language models, and explainability that are unlikely to be addressed elsewhere. As such, instead of classification, we turn our attention to sentence encodings for similarity search, which are also part of one of the case studies. Another interesting subject is to determine topic evolution over time.

The most delicate conversation is undoubtedly about large language models. We view this as an opportunity to debate recurring issues in today's society regarding privacy and ownership. As previously mentioned, completely avoiding new technologies is unfeasible. Our conversations will revolve around local LLMs and efficient inference on consumer hardware, explaining the surge of mixture of experts models.

Due to practical considerations, such as needing to synchronize with other courses and teachers across several study programs, there is limited flexibility in what can be changed in the covered topics themselves.

### 3.2 Case Studies

To encourage students to get involved in research activities, we discuss how we have successfully applied "old-school" methods in modern settings.

1. Our raw data consisted of textbooks in PDF format that had selectable and unobfuscated text. We employed two methods to obtain the

desired content: manual parsing with dedicated libraries and LLM extraction. We deduplicate the extracted sentences, noting that each method managed to obtain samples not found by the other method. Next, we use POS tagging and dependency parsing for frequency analyses to spot certain trends.

2. In a food safety context where gold annotations were noisy, we reformulated the task as an information retrieval problem and applied sequence similarity. Since this only requires content words, we remove stop words and apply lemmatization. We wanted to return the results ordered by importance, and this is not something that can be defined algorithmically. Thus, the final results are obtained with reranking by a custom sorting algorithm that had a LLM call as a comparison function.

3. Two cascaded NER models with clever post-processing outperformed LLMs on certain tasks of relation extraction (Farzi et al., 2024), revealing a difference of orders of magnitude in terms of computational costs. When constrained by lack of resources, developing efficient solutions is a must. Conversely, when having access to AI accelerators and other expensive hardware, there is little incentive to optimize these costs.

These examples show that valuable results can be achieved with modest resources. We observe an increasing number of undergraduate students that publish papers as part of courses held by our group: from 2-4 students with 2-4 papers in 2023 and 2024 to 19 students with 12 papers in 2025.[3] However, it can also be attributed to the growing interest in this field.

## 4 Student Evaluation

Evaluation consists of a team project of 2 to 4 people that is presented at the end of the course, together with a technical report in the form of a research paper. Students must choose between a literature survey and an end-to-end application that tackles a real problem. Alternatively, they are encouraged to participate in relevant competitions and shared tasks.

Since our goal is to cultivate critical thinking and a sense of responsibility, we emphasize ablation studies and detailed error analyses, as opposed to simply displaying scores, a practice unfortunately very common in the field. We share our vision, teaching methodologies, and evaluation criteria in the hope that the community will benefit from these insights.

## 5 Conclusion

We have presented our approach for an introductory course in natural language processing, providing a middle ground solution for balancing fundamental and classical techniques with the pervasiveness of large language models. We underline that misuse of these tools is solely the result of lack of ethical and moral grounding in society itself, which can be remedied by education. As long as humans always stay in the loop and understand every part of the process, including its limitations, we think that adequate adoption will be beneficial.

Regarding LLM recommendations to students, we would like to promote open-weight and open-source local models, since they offer more control and complete privacy. Other reasons include better reproducibility and availability because commercial LLMs are often subject to rate limits, even on paid plans.

Our teaching materials are publicly available on GitHub,[4] in line with Open Source Educational Resources principles (Bothmann et al., 2023).

## Limitations

Teaching materials are living documents that require constant revision. We do not claim to have found the perfect formula neither in terms of educational resources, nor in pedagogical approaches. We believe that open exchange of knowledge and experience is the most effective path to achieve quality education.

## Ethical Considerations

I have used free tiers of commercial LLMs for parts of literature review and a mix of commercial and local LLMs for coding assistance. No LLMs were used in the writing of this paper.

To the best of my knowledge, there is very little attention given to local models compared to the commercial counterparts in discourses related to LLM use in education. This is unfortunate because local models offer a great starting point to discuss problems of ownership and privacy.

[3] See for example Sergiu Nisioi and Ana Sabina Uban (student coordinators) on ACL Anthology.

[4] https://github.com/mcmarius/inlp

While there are many valid arguments raised by Bender and Hanna (2025) that I agree with, including the direction of AI research itself, I have to note that extreme attitudes in both directions show a lack of understanding of the subject. As stated before, the issue of misapplication of LLMs is the fault of the stakeholders taking such decisions, not of the tools themselves. Those with ill intentions will find harmful ways with or without LLMs, local or commercial. The negative position mentioned here only serves to promote commercial models even more, further marginalizing contributions on local LLMs, which are not referenced at all in that work. Ignoring these technologies for those actively involved in the field is no longer realistic at this point. We should look into solving the root causes, not the symptoms.

Beside my academic duties, I am also working in the industry, making use of and developing LLM tooling. This opportunity helped me understand that sensible usage of language models provides actual value, beyond all the hype in the media (both positive and negative). I am aware that my views of LLM usage and evolution might seem overly optimistic given this context. Even though I am using commercial LLMs for professional purposes, my personal stance is towards local LLMs as much as possible.

## Acknowledgments

This research is supported by the project "Romanian Hub for Artificial Intelligence - HRIA", Smart Growth, Digitization and Financial Instruments Program, 2021-2027, MySMIS no. 351416.

We would like to thank the reviewers for their thoughtful, rigorous, and constructive feedback, which has helped sharpen our arguments, clarify key points, and improve the overall quality of this work.

## References

Matthias Aßenmacher, Andreas Stephan, Leonie Weissweiler, Erion Çano, Ingo Ziegler, Marwin Härttrich, Bernd Bischl, Benjamin Roth, Christian Heumann, and Hinrich Schütze. 2024. Collaborative development of modular open source educational resources for natural language processing. In *Proceedings of the Sixth Workshop on Teaching NLP*, pages 43–53, Bangkok, Thailand. Association for Computational Linguistics.

Emily M. Bender and Alex Hanna. 2025. *The AI Con: How to Fight Big Tech's Hype and Create the Future We Want*. HarperCollins, New York.

Ludwig Bothmann, Sven Strickroth, Giuseppe Casalicchio, David Rügamer, Marius Lindauer, Fabian Scheipl, and Bernd Bischl. 2023. Developing open source educational resources for machine learning and data science. In *Proceedings of the Third Teaching Machine Learning and Artificial Intelligence Workshop*, volume 207 of *Proceedings of Machine Learning Research*, pages 1–6. PMLR.

Randall Curren. 2025. Philosophy of Education. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*, Spring 2025 edition. Metaphysics Research Lab, Stanford University.

Saeed Farzi, Soumitra Ghosh, Alberto Lavelli, and Bernardo Magnini. 2024. Get the best out of 1B LLMs: Insights from information extraction on clinical documents. In *Proceedings of the 23rd Workshop on Biomedical Natural Language Processing*, pages 266–276, Bangkok, Thailand. Association for Computational Linguistics.

Annemarie Friedrich and Torsten Zesch. 2021. A crash course on ethics for natural language processing. In *Proceedings of the Fifth Workshop on Teaching NLP*, pages 49–51, Online. Association for Computational Linguistics.

Aditya Joshi, Jake Renzella, Pushpak Bhattacharyya, Saurav Jha, and Xiangyu Zhang. 2024. Striking a balance between classical and deep learning approaches in natural language processing pedagogy. In *Proceedings of the Sixth Workshop on Teaching NLP*, pages 23–32, Bangkok, Thailand. Association for Computational Linguistics.

Daniel Jurafsky and James H. Martin. 2025. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, with Language Models*, 3rd edition. Online manuscript released August 24, 2025.

Viggo Kann. 2025. Students' attitudes towards cheating before and after ChatGPT. In *Proceedings of the 30th ACM Conference on Innovation and Technology in Computer Science Education V. 1*, ITiCSE 2025, page 291–297, New York, NY, USA. Association for Computing Machinery.

Roger McDermott, Mats Daniels, John N.A. Brown, and Åsa Cajander. 2025. Determining the scope of the philosophy of computing education. In *Proceedings of the 30th ACM Conference on Innovation and Technology in Computer Science Education V. 1*, ITiCSE 2025, page 403–409, New York, NY, USA. Association for Computing Machinery.

Shotaro Naganuma, Tsubasa Minematsu, Sachika Shibukawa, Aikana Ohno, and Yukinori Wakihama. 2025. Exploring students' generative AI usage patterns and knowledge creation in collaborative problem solving. In *Artificial Intelligence in Education*.

*Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium, Blue Sky, and WideAIED*, pages 305–312, Cham. Springer Nature Switzerland.

Irina Nikishina, Maria Tikhonova, Viktoriia Chekalina, Alexey Zaytsev, Artem Vazhentsev, and Alexander Panchenko. 2024. Industry vs academia: Running a course on transformers in two setups. In *Proceedings of the Sixth Workshop on Teaching NLP*, pages 7–22, Bangkok, Thailand. Association for Computational Linguistics.

Baichang Zhong and Xiaofan Liu. 2025. Evaluating AI literacy of secondary students: Framework and scale development. *Computers & Education*, 227:105230.