Teaching NLP in the AI Era: Experiences from the University of Latvia

Inguna Skadiņa and **Guntis Bārzdiņš** and **Uldis Bojārs** and
Normunds Grūzītis and **Pēteris Paikens**
University of Latvia, Faculty of Science and Technology
Raiņa bulv. 19, Riga, Latvia
firstname.lastname@lu.lv

Abstract

From being a niche technology with practical applications in translation and speech recognition, NLP is now underpinning the AI era through LLMs, promising a universal economic impact in the future. Although transitioning to the AI era is hyped by BigTech companies, practical adoption of the LLM capabilities for economically impactful tasks and processes goes via education of specialists capable to apply it properly. Human-in-the-loop, accuracy measurement, fine-tuning, on-premises processing of sensitive data have become essential skills for applying NLP. This short paper introduces two language technology modules developed and piloted at the Faculty of Science and Technology of the University of Latvia.

1 Introduction

Although various computational linguistics (CL) and natural language processing (NLP) courses have been offered at universities in Latvia for almost two decades, the rapid progress in the application of artificial intelligence (AI) and the continued advancement of language technology (LT) have substantially increased the demand for LT study modules and specialized courses.

The need for designated LT study modules has been recognized by the Ministry of Education and Sciences of Latvia, proposing to invest in the development of high-level digital skills to significantly increase the number of LT specialists by 2026. The Language Technology Initiative (LTI) project[1] was initiated to implement this objective through synergy between the research and higher education sectors. The project aims to develop a curriculum for cross-disciplinary LT teaching, advance Latvian language resources, and create platforms and tools for studying and experimentation, as well as for conducting research by young researchers (Skadiņa et al., 2024).

We present two LT modules for computer science (CS) students, developed through the LTI project at the Faculty of Science and Technology of the University of Latvia (UL). Overview of courses. their timing and credit weight is summarized in Table 1.

While the developed modules and courses are grounded in widely used teaching materials, they also integrate Latvian-specific content and showcase Language Technology solutions developed and tailored for the less-resourced language Latvian. This dual approach ensures that students gain a solid foundation in established natural language processing methods, while simultaneously engaging with practical challenges, language resources, and tools relevant to the Latvian language. By combining international best practices with local language resources, the courses not only support the development of technical skills but also foster awareness of linguistic diversity and the challenges of working with less-resourced languages.

The teaching materials of both modules are available to students via an e-studies platform of UL.[3] In addition to traditional slide-based lectures, all courses make systematic use of Jupyter notebooks, which are freely accessible to students as well as external users.

2 Bachelor's Module for CS Students

The bachelor's module in LT comprises three courses: Fundamentals of Natural Language Processing, Fundamentals of Deep Machine Learning, and Introduction to Python Programming Language.

The Python course establishes a foundation in Python programming that other courses can build upon. It covers both introductory and more advanced topics such as text processing, data analysis, and visualization. To support interactive and

---

[1] https://www.vti.lu.lv/en/

[3] https://estudijas.lu.lv

Course	Module	Semester	ECTS	Language
Fundamentals of Deep Machine Learning	Bachelor	4	3	Latvian
Introduction to Python Programming Language	Bachelor	5	3	Latvian
Fundamentals of Natural Language Processing	Bachelor	6	3	Latvian, English
Web Information Retrieval Techniques[2]	Bachelor	6	3	English
Selected Topics in Deep Learning[1]	Bachelor	6	3	Latvian
Applications of Language Technology	Master	1	3	Latvian
Deep Machine Learning	Master	2	6	Latvian

Table 1: Overview of Language Technology courses and modules at the Faculty of Science and Technology

practice-oriented learning, we have developed a set of Jupyter notebooks.[4] By interactively using the notebooks, students can better follow the material presented by the instructor (Granger and Pérez, 2021).

The Deep Machine Learning course provides an overview of modern applications of machine learning and develops practical skills in using deep neural networks for common machine learning tasks, including NLP. Initially, students are taught the fundamental machine learning concepts with a focus on neural networks, and then moving on to practical applications using the PyTorch framework, covering the implementation details and training of widely applied types of neural networks – perceptron classifiers, convolution and recurrent networks, and transformer methods. This is a hands-on course with the key learning outcomes being tied to practical projects in classification, text analysis, and image processing. It also covers software engineering challenges regarding training and deployment of models, as well as working with existing third-party models and embeddings, and fine-tuning pre-trained language models for specific tasks, which is now a common approach in NLP.

The Natural Language Processing course introduces students to fundamental CL and LT problems and methods, as well as practical applications, comparing the current capabilities for high-resource and low-resource languages. The course covers core methods as well as key innovations and trends in the LT field, including for Latvian. It encompasses language processing and modeling at various levels of text analysis, utilizing both knowledge-based and data-driven approaches. The main focus is on data-driven methods and the language resources they require. The course is partially based on selected chapters from the Jurafsky and Martin (2025) text-book and some mate-

rials from (Nugues, 2024), adapted and extended for the Latvian context. The Jupyter notebooks used in this course are available from a GitHub repository.[5] Language resources and tools available from the CLARIN-LV repository,[6] as well as those developed by the Institute of Mathematics and Computer Science, UL (the research partner of the LTI project), are integrated into the course to introduce students to state-of-the-art language technology resources and tools for Latvian. During the course, students are required to complete two assignments: one focused on rule-based methods (finite-state automata for morphological analysis), and another centered on data-driven methods (classification). To successfully complete the course, for exam students (usually working in pairs) must develop a small practical NLP solution, perform an evaluation, demonstrate its functionality, and submit comprehensive documentation. This hands-on approach allows students to apply both theoretical knowledge and computational techniques to real-world language processing tasks, fostering problem-solving and critical thinking skills.

3 Master's Module for CS Students

The master's module in LT comprises two fundamental courses: Applications of Language Technology and Deep Machine Learning.

The applied Language Technology course introduces state-of-the-art Transformer-based approaches for language understanding and generation, including speech recognition and synthesis. It provides the necessary theoretical background while focusing on practical skills, allowing students to understand, apply and integrate, or even fine-tune large language models (LLM) for practical applications. To acquire these skills, open-source LLMs, datasets, and programming frameworks are

---

[4]https://github.com/CaptSolo/LU_Python_course

[5]https://github.com/LUMII-AILab/NLP_Course
[6]https://repository.clarin.lv

used.[7]

To address the teaching challenges posed by the rapid progress of LT, the course includes invited lectures and tutorials delivered by external experts, also from the industry. The aim of this initiative is twofold: to familiarize students with practical problems and state-of-the-art LT solutions, and to provide instructors with new teaching materials that can be adopted and reused in future iterations of the course.

After the first run of the LT course in Fall 2025, overall student feedback has been very positive. Some students even highlighted it as the most modern and interesting course of the semester. Knowledge transfer from national and international research and industry experience was also acknowledged by students. The hands-on approach proved to be more time-consuming than initially expected by both instructors and students; however, students admitted that it helped them better understand the theory and acquire practical skills that would otherwise have remained purely theoretical. The first iteration of the course also revealed that while students are familiar with decoder models as black boxes, they require a slower-paced introduction to the underlying concepts of the Transformer architecture and neural language models in general, as well as to the encoder models in particular.

The Deep Learning course introduces students to the currently popular deep machine learning methods. As GPUs become more powerful and the available datasets larger, deep neural networks have gradually replaced simpler machine learning methods in scenarios that can leverage pre-training on large quantities of unlabeled data, which includes many NLP tasks. During this course, students get acquainted with the concepts of supervised, unsupervised, and self-supervised machine learning from primitive classification to large language models and generative AI. Reinforcement and evolutionary learning for robotics is also discussed. Benchmarking against human performance is used to explain the great challenges and the limited economic impact of current models. Jupyter notebooks for this course are available on GitHub.[8]

4   Conclusion and Future Plans

We provided a concise overview of two language technology modules developed for computer science students at the University of Latvia. More information about the modules and courses is available on the Language Technology Initiative website.[9] Two additional bachelor-level courses – Selected Topics in Deep Learning, and Web Information Retrieval Techniques – are included in the curriculum for the Spring semester of 2026.

During the Autumn semester of 2025, several courses from these LT modules were offered to minor–major students from other faculties and programs. This initiative aimed to broaden access to LT education beyond computer science and to support interdisciplinary learning. Seven students from linguistics, history, and geography were enrolled into the courses and successfully completed the minor–major module.

Acknowledgments

References

Brian E. Granger and Fernando Pérez. 2021. Jupyter: Thinking and storytelling with code and data. *Computing in Science & Engineering*, 23(2):7–14.

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.

Pierre M. Nugues. 2024. *Python for Natural Language Processing Programming with NumPy, scikit-learn, Keras, and PyTorch*. Springer Nature, Switzerland.

Inguna Skadiņa, Jana Kuzmina, Sergejs Kruks, Marina Platonova, Tatjana Smirnova, and Ilze Auzina. 2024. Language Technology Initiative - Bridging the Gap between Research and Education. In *CLARIN Annual Conference Proceedings*, pages 26–30.

---

[7]https://github.com/LUMII-AILab/NLP_Course
[8]https://github.com/guntisx/DeepLearningCourse
[9]https://www.vti.lu.lv/en/education/study-modules/