



# Nikolay Paev

✉ Email: [nikolay.paev@iict.bas.bg](mailto:nikolay.paev@iict.bas.bg)

🌐 LinkedIn: [linkedin.com/in/nikolay-paev-03b566292](https://www.linkedin.com/in/nikolay-paev-03b566292)

## WORK EXPERIENCE

### *Institute of Information and Communication Technologies - Bulgarian Academy of Sciences*

[ 01/02/2026 – Current ] **Assistant researcher in higher education**

- Training Large Language Models for Bulgarian from scratch (pre-training).
- Experience in working with the High-performance computational cluster HEMUS.
- Developing and training (fine-tuning) language models for different tasks in the fields of Natural Language Processing and Information Extraction.
- Writing scientific publications and presenting works at conferences.

### *Faculty of Mathematics and Informatics - Sofia University "St. Kliment Ohridski"*

City: Sofia | Country: Bulgaria

[ 01/10/2024 – 25/02/2025 ] **University teaching assistant**

- Teaching seminars in "Knowledge modelling and representation" to first year Master students in Artificial Intelligence.
- Grading course projects.

### *Institute of Information and Communication Technologies - Bulgarian Academy of Sciences*

City: Sofia | Country: Bulgaria

[ 01/12/2023 – 01/02/2026 ] **Software developer**

## EDUCATION AND TRAINING

[ 01/10/2023 – 07/07/2025 ] **Master of Science in Informatics - Artificial Intelligence**

*FMI at Sofia University "St. Kliment Ohridski"* <https://www.fmi.uni-sofia.bg/en/artificial-intelligence>

City: Sofia | Country: Bulgaria | Field(s) of study: Information and Communication Technologies | Final grade: 6.0/6.0 | Level in EQF: EQF level 7 | Thesis: Automatic extraction of questions from exams

Theoretical foundations and practical knowledge in artificial intelligence, machine learning, knowledge representation, information retrieval, natural language processing, semantic technologies, and recommender systems.

[ 01/10/2019 – 12/07/2023 ] **Bachelor in Computer Science**

*FMI at Sofia University "St. Kliment Ohridski"* <https://www.fmi.uni-sofia.bg/en/computer-science>

City: Sofia | Country: Bulgaria | Field(s) of study: Information and Communication Technologies | Final grade: 5.72/6.00 | Level in EQF: EQF level 6

Strong theoretical and practical skills in mathematics, discrete and computational theories, computer systems and architectures, programming paradigms, databases, algorithms, and software development.

## SKILLS

---

natural language processing | machine learning | write scientific publications | deep learning | artificial neural networks | Python (computer programming) | Python & libraries (Numpy, Matplotlib, Scikit-learn, Pandas, Pytorch) | assist students in their learning

## PROJECTS

---

- [ 01/12/2023 – Current ] **Bulgarian national research infrastructure for resources and technologies for linguistic, cultural and historical heritage, integrated within CLARIN EU and DARIAH EU (CLaDA-BG)**
- [ 01/09/2025 – Current ] **Exploring the Triad of Academic Writing, Critical Thinking and AI Literacy: Interdisciplinary Perspectives on Undergraduate Thesis Writing (TRAI)**
- [ 01/05/2025 – Current ] **ParlaCAP - Comparing agenda settings across parliaments via the ParlaMint dataset**
- [ 01/12/2025 – Current ] **Digital transformation in organizations: Model of key psychological determinants in human – artificial intelligence interaction**

## LANGUAGE SKILLS

---

**Mother tongue(s):** Bulgarian

**Other language(s):**

**English**

**LISTENING C1 READING C1 WRITING C1**

**SPOKEN PRODUCTION C1 SPOKEN INTERACTION B2**

*Levels: A1 and A2: Basic user; B1 and B2: Independent user; C1 and C2: Proficient user*

## PUBLICATIONS

---

- [ 2025 ] [\*\*Word Sense Disambiguation with Large Language Models: Casing Bulgarian\*\*](#)
- Reference:** Paev, Nikolay and Simov, Kiril and Osenova, Petya “Word Sense Disambiguation with Large Language Models: Casing Bulgarian”, Proceedings of the 13th International Global Wordnet Conference (GWC2025) Pages 171–178, Pavia, 2025  
**Authors:** Nikolay Paev and Kiril Simov and Petya Osenova | **Journal Name:** 13th International Global Wordnet Conference (GWC2025) | **Volume, Issue and Pages:** 171-178
- [ 2024 ] [\*\*Introducing Shallow Syntactic Information within the Graph-based Dependency Parsing\*\*](#)
- Reference:** Nikolay Paev and Kiril Simov and Petya Osenova, “Introducing Shallow Syntactic Information within the Graph-based Dependency Parsing”, Proceedings of the 22nd Workshop on Treebanks and Linguistic Theories (TLT 2024), Pages 46--54, Hamburg, 2024  
**Authors:** Nikolay Paev and Kiril Simov and Petya Osenova | **Journal Name:** 22nd Workshop on Treebanks and Linguistic Theories (TLT 2024) | **Volume, Issue and Pages:** 46-54
- [ 2025 ] [\*\*Bulgarian Event Extraction with LLMs\*\*](#)
- Reference:** Kiril Simov, Nikolay Paev, Petya Osenova, Stefan Marinov “Bulgarian Event Extraction with LLMs”, Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing (RANLP2025), Pages 1163-1171, Varna, 2025  
**Authors:** Kiril Simov, Nikolay Paev, Petya Osenova, Stefan Marinov | **Journal Name:** 14th International Conference on Recent Advances in Natural Language Processing (RANLP2025) | **Volume, Issue and Pages:** 1163-1171

[ 2025 ] **LLM-based Models for Transforming of Diachronic Bulgarian Spelling to Contemporary Bulgarian**

**Reference:** Kiril Simov, Nikolay Paev and Petya Osenova, "LLM-based Models for Transforming of Diachronic Bulgarian Spelling to Contemporary Bulgarian", Proceedings of CLARIN Annual Conference 2025 (CLARIN2025), Vienna, 2025

**Authors:** Kiril Simov, Nikolay Paev, Petya Osenova | **Journal Name:** CLARIN Annual Conference 2025 (CLARIN2025)

[ 2025 ] **An Event-centered Knowledge Graph for Bulgaria**

**Reference:** Kiril Simov, Nikolay Paev and Petya Osenova, "An Event-centered Knowledge Graph for Bulgaria", Proceedings of 1st GOBLIN Workshop on Knowledge Graph Technologies (GOBLIN2025), Leipzig, 2025

**Authors:** Kiril Simov, Nikolay Paev and Petya Osenova | **Journal Name:** 1st GOBLIN Workshop on Knowledge Graph Technologies (GOBLIN2025)

[ 2025 ] **Visualization of LLM Annotated Documents**

**Reference:** Teodor Valchev, Nikolay Paev, "Visualization of LLM Annotated Documents", Proceedings of the Student Research Workshop at 14th International Conference on Recent Advances in Natural Language Processing (RANLP-SRW2025), Varna, 2025

**Authors:** Teodor Valchev, Nikolay Paev | **Journal Name:** Student Research Workshop at 14th International Conference on Recent Advances in NLP(RANLP-SRW2025)

## DRIVING LICENCE

---

**Cars:** B