

Michal Štefánik – Curriculum Vitae

Personal information

Personal website: <https://michal-stefanik.github.io>
Personal email address: stefanik.mich@gmail.com
Social account (LinkedIn): <https://linkedin.com/in/stefanikm>
Date of birth: 12/12/1994

Professional experience

Oct 2025 – now	Specially Appointed Researcher	R&D Center for Large Language Models, National Institute of Informatics (NII), Japan Advised by Pontus Stenetorp (UCL/NII)
Feb 2025 – Oct 2025	Postdoctoral Research Associate	Language Technology group, University of Helsinki Advised by Jörg Tiedemann
Sep 2021 – now	AI Researcher in Education*	Faculty of Education, Masaryk University
Feb 2024 – May 2024	Visiting Researcher: Modularization	Language Technology group, University of Helsinki
Mar 2021 – Jan 2025	NLP Researcher & Team Lead*	Gauss Algorithmic, s.r.o
Jun 2018 – Mar 2021	Deep Learning Engineer*	Gauss Algorithmic, s.r.o
Mar 2015 – Jun 2018	Software Developer*	Red Hat Inc.

*part-time positions

Education

Sep 2016 – Feb 2020	Ph.D, Fundamentals of Computer Science Ph.D Dissertation title: <i>Axes of Robustness of Neural Language Models</i> defended <i>Summa Cum Laude</i> ¹ . Ph.D award date: 06/02/2025	Faculty of Informatics, Masaryk University
Sep 2016 – Feb 2020	Master of Science, Artificial Intelligence and Natural Language Processing	Faculty of Informatics, Masaryk University
Sep 2013 – Jun 2015	Bachelor of Science, Applied Informatics	Faculty of Informatics, Masaryk University

Peer-reviewed publications

1. F. Sadrieh, M. Štefánik: **Prompt-Robust Language Models: Which Training Strategies Work?** In Findings of EMNLP 2026 (CORE A* conference, acceptance 29.7%).
2. M. Štefánik, P. Mondorf, A. Waldis, Q. Liu, C. Yang, M. Spiegel, J. Kuchař, M. Kadlčík, AV. Oomerjee, C. Liu, S. Frieder, B. Plank, F. Barez, P. Stenetorp: **AIMO Interpretability Challenge**. NeurIPS 2026 Competitions track (CORE A* conference, acceptance 14.1%).
3. M. Štefánik, T. Mickus, M. Kadlčík, B. Højer, M. Spiegel, R. Vazquez, A. Sinha, J. Kuchař, P. Mondorf, P. Stenetorp: **Language Models Learn Universal Representations of Numbers and Here's Why You Should Care**. In Proceedings of ACL 2026 (CORE A* conference, acceptance 18.9%) & Spotlight at ICML 2026 Mechanistic Interpretability workshop (acceptance 2.8%).
4. M. Štefánik, T. Mickus, M. Spiegel, M. Kadlčík, J. Kuchař: **Can Out-of-Distribution Evaluations Uncover Reliance on Prediction Shortcuts? A Case Study in Question Answering**. In Findings of EMNLP 2025 (CORE A* conference, acceptance 39.51%).
5. M. Kadlčík*, M. Štefánik*, T. Mickus*, J. Kuchař, M. Spiegel (*core): **Pre-trained Language Models Learn Remarkably Accurate Representations of Numbers**. In Proceedings of EMNLP 2025 (CORE A* conference, acceptance 22.16%).
6. T. Vrabcová, M. Kadlčík, P. Sojka, M. Štefánik, M. Spiegel: **Negation: A Pink Elephant in the Large Language Models' Room?** In Findings of EMNLP 2025 (CORE A* conference, acceptance 39.51%).
7. M. Fajčík, M Docekal, J Dolezal, K Ondrej, K Beneš, J. Kapsa, P. Smrz, A. Polok, M. Hradis, Z. Neverilova, A. Horak, R. Sabol, M. Štefánik, A. Jirkovsky, D. Adamczyk, P. Hyner, J. Hula, H. Kydlíček: **BenCzechMark: A Czech-centric Multitask and Multimetric Benchmark for Large Language Models with Duel Scoring Mechanism**. In Transactions of ACL (2024 IF 6.9)

(1) Link to the dissertation thesis: <https://is.muni.cz/th/m805b/?lang=en>

8. M. Kadlčík*, M. Štefánek* (*equal): **Self-Training Language Models in Arithmetic Reasoning**. In Findings of EMNLP 2024 (CORE A* conference, acceptance 37.7%).
9. M. Štefánek, M. Kadlčík, P. Sojka: **Concept-aware Data Construction Improves In-context Learning of Language Models**. In Findings of Annual Meeting of ACL (CORE A* conference, acceptance 43.4%).
10. L. Mikula*, M. Štefánek*, M. Petrovič, P. Sojka (*equal): **Think Twice: Measuring the Efficiency of Eliminating Prediction Shortcuts of Question Answering Models**. In Proceedings of EACL (CORE A conference, acceptance 20.3%).
11. M. Kadlčík*, M. Štefánek*, O. Sotolář, V. Martinek (*equal): **Calc-X and Calcformers: Empowering Arithmetical Chain-of-Thought through Interaction with Symbolic Systems**. In Proceedings of EMNLP 2023 (CORE A* conference, acceptance 21.3%).
12. M. Štefánek, M. Kadlčík, P. Sojka: **Soft Alignment Objectives for Robust Adaptation of Language Generation**. In Proceedings of Annual Meeting of ACL 2023 (CORE A* conference, acceptance 23.5%).
13. M. Štefánek, M. Kadlčík: **Can In-Context Learners Learn a Reasoning Concept from Demonstrations?** In Proceedings of ACL 2023 Natural Language Reasoning and Structured Explanations (Best Paper award).
14. J. Nehyba*, M. Štefánek* (*equal): **Applications of deep language models for reflective writings**. In Education and Information Technologies (Q1 in Education, 2024 Impact Factor 5.4, acceptance ~15%).
15. M. Štefánek, M. Kadlčík, P. Sojka: **Resources and Few-shot Learners for In-context Learning in Slavic Languages**. In Proceedings of EACL 2023 SlavicNLP (Best Paper award).
16. M. Štefánek: **Methods for Estimating and Improving Robustness of Language Models**. In Proceedings of NAACL Student Research Workshop 2022 (Thesis proposal, acceptance 47%).
17. M. Štefánek, V. Novotný, N. Groverová, P. Sojka: **Adaptor: Objective-Centric Adaptation Framework for Language Models**. In Proceedings of ACL 2022: Demonstrations (CORE A* conference, acceptance 36%).
18. M. Štefánek, V. Novotný, P. Sojka: **Regressive Ensemble for Machine Translation Quality Evaluation**. In Proceedings of WMT 2021 (GScholar #9 in Computational Linguistics).

Selected talks

1. **Uncertainty and Understanding as Foundations for Robust Language Models**. Invited talk in MainLP Lab at Ludwig Maximilian University of Munich. 08/04/2026, Munich, Germany.
2. **Robust Language Models One Step at a Time**. Invited talk at Kempelen Institute of Information Technology. 27/06/2025, Bratislava, Slovakia. Supported by HORIZON-CSA Grant 101136646.
3. **Why We Should Care about Uncertainty when Training Language Models**. Invited talk at Brno University of Technology. 14/04/2025, Brno, Czechia.
4. **Reliable Language Models and Where to Find Them**. Invited talk at the Faculty of Humanities, University of Helsinki. 13/03/2025, Helsinki, Finland.
5. **Fantastically Robust Language Models and Where to Find Them**. Invited talk at Announcement of Outstanding PhD Researcher at Faculty of Informatics. 29/10/2024, Brno, Czechia.
6. **Fantastically Robust Language Models and Where to Find Them**. Industrial outreach talk at Allegro Research Seminar 09/10/2024, Warsaw, Poland.
7. **Language Models in Autonomous Systems**. Invited popularisation talk at DataMesh community meetup. 12/09/2024 Brno, Czechia.
8. **Training for Single Correct Prediction Makes Your Model More Fragile**. Scientific popularisation talks at Human-Aligned Summer School, 18/07/2024, Prague, Czechia, and ELLIS Robust LLMs Workshop, 30/07/2024, Oxford, United Kingdom.
9. **Think Twice: Measuring the Efficiency of Eliminating Prediction Shortcuts of QA Models**. Oral presentation at EACL conference (CORE A). St. Julian's, Malta, 19/03/2024.
10. **Robustness of Language Models and Perspectives of Modularization**. Invited talk at Faculty of Humanities, University of Helsinki. 07/03/2024, Helsinki, Finland.
11. **Can In-context Learners Learn a Reasoning Concept from Demonstrations?** Invited talk accompanying Best Paper Announcement, ACL Natural Language Reasoning and Structured Explanations workshop, 13/07/2023, Toronto, Canada.
12. **Learning to Learn: Hands-on Tutorial in Using and Improving Few-shot Language Models**. Accepted industrial tutorial with presentation at ML Prague conference, 02/06/2023, Prague, Czechia.
13. **Resources and Few-shot Learners for In-context Learning in Slavic Languages**. Invited talk accompanying Best Paper Announcement, EACL SlavicNLP workshop. 01/05/2023, Dubrovnik, Croatia.
14. **Learning to Learn: Training Language Models to Understand Tasks from Few Examples**. Invited popularisation talk at ML Meetups community meetup. 03/10/2022, Brno, Czechia

15. **Mitigating Biases of QA Models by Simple Resampling Methods.** Invited talk accompanying Announcement of winning submission at NAACL Dynamic Adversarial Data Collection 2022. Seattle, US.
16. **Methods for Estimating and Improving Robustness of Language Models.** Oral presentation at NAACL Student Research Workshop 2022. Seattle, US

Software and Data

1. **Adaptor: Objective-Centric Adaptation Framework for Language Models:** Open-source software library¹ enabling easier training of language models using advanced robustness-enhancing methods. Adaptor exemplifies **successful industrial innovation** — accelerating the productisation and commercialisation of most advanced research outcomes thanks to innovative software design. As of 01/08/2025, Adaptor-trained models were used by over 8,000 monthly users.²
2. **Calformers:** First freely available implementation of tool-using language models³ — fostered and accelerated the emergence of the now established subfield of NLP (Agentic and Tool-using LMs). I am an equal first coauthor of this repository.
3. **Calc-X:** Unified data collection supporting the training of state-of-the-art mathematical reasoning models. As of 01/08/2025, used by over 1,000 researchers. Collection is freely available under open licenses.⁴ I am an equal first coauthor of this repository.
4. **CEReD & Edustories:** Open datasets, models, and implementations⁵ resulting from my efforts in opening Educational research. We are aware of these resources being employed by at least three groups at EPFL, MU, and CMU. I am the first author of this repository.
5. **BenCzechMark:** A comprehensive collection of language resources⁶ for Czech, created within the Czech Language Consortium (CZLC), involving all major Czech universities. I am a co-founder and promoter of the CZLC initiative and a designer of evaluation protocols deployed in BenCzechMark.

Mentoring and Supervision

- Founder and leader of **TransformersClub** — a collaborative platform supporting students interested in NLP/AI research. Current and past members I supervise(d), and a summary of their achievements:

Jun 2025 – now	Frederic Sadrieh	First-authored paper at EMNLP 2026
Sep 2024 – now	Michal Spiegel	First-authored paper (under review) ⁷
Sep 2024 – now	Josef Kuchař	First-authored LREC 2026 paper ⁸ , Scientific dataset ⁹
Sep 2022 – now	Marek Kadlčík	First-authored papers: EMNLP 2023, EMNLP 2024, EMNLP 2025 Dean's Award (~5% of theses), Rector's Award ¹⁰ (~0.1% of MSc students)
Sep 2022 – Feb 2024	Šárka Ščavnická	Scientific dataset ¹¹
Sep 2022 – Feb 2024	Dávid Meluš	Open-source models ¹²
Sep 2021 – Feb 2023	Lukáš Mikula	First-authored EACL 2024 paper, Dean's Award
Sep 2021 – Jun 2022	Marek Petrovič	3rd place in International SVOC Competition ¹³ , Dean's Award
Feb 2022 – Jun 2022	Matej Meško	Open-source contributions
Feb 2020 – Jun 2021	Martin Geletka	Nomination for a Dean's Award
Sep 2020 – Jun 2021	Petr Mička	Improving quality of teaching materials

(1) Software repository with Adaptor implementation: <https://github.com/gaussalgo/adaptor>;

(2) Based on statistics of known models available on HuggingFace repository: <https://huggingface.co/models>;

(3) Software repository with Calformers implementation: <https://github.com/prompTEUS/calc-x>;

(4) Freely-available Calc-X collection: <https://huggingface.co/collections/MU-NLPC/calc-x-652fee9a6b838fd820055483>;

(5) All CEReD & Edustories are freely available on HuggingFace repository:

<https://huggingface.co/collections/MU-NLPC/edustories-66e5ed0a6998c3d86cd10a93>

(6) All BenCzechMark data is freely available on HuggingFace repository: <https://huggingface.co/CZLC/datasets>

(7) M. Spiegel et al.: Attend or Perish: Benchmarking Attention in Algorithmic Reasoning. arXiv:2503.01909

(8) J. Kuchař et al.: VectorEdits: A Dataset and Benchmark for Instruction-Based Editing of Vector Graphics. arXiv:2506.15903

(9) Dataset repository: <https://huggingface.co/datasets/mikronai/VectorEdits>

(10) News article: <https://www.em.muni.cz/en/news/18093-dies-academicus-2025-was-held-in-spirit-of-ideas-of-universitas>

(11) CIVQA datasets at: <https://huggingface.co/datasets/fimu-docproc-research>

(12) Language models published on <https://huggingface.co/fimu-docproc-research/models>

(13) Announcement repository (in the language of Organisers): <https://www.math.sk/svoc2022/#vysledky>

Committee memberships

- **Member of Program Committees** at *ACL, NeurIPS, and satellite workshops:
 - **Reviewer of 30+ papers:** EACL 2023, ACL 2023, EMNLP 2023, EACL 2024, ACL 2024, GenBench workshop at EMNLP 2024, EMNLP 2024, BabyLM workshop at EMNLP 2025, EMNLP 2026
 - **Area Chair of 50+ papers:** NeurIPS 2026 LLMs track, ACL 2026 Interpretability & Explainability tracks, EMNLP 2025 Low-resource track, ACL 2025 Generalization track
- **Main organiser and Program Chair** at NeurIPS 2026 AIMO Interpretability Challenge

Prior funding

Jul 2026 – Dec 2026	Grant name: Awards for the NeurIPS 2026 AIMO Interpretability Challenge Granted funding: 11,662 EUR, source: XTX Markets Limited, London, UK Role: Principal Investigator
Feb 2025 – Sep 2026	Grant name: GreenNLP – controlling the carbon footprint in sustainable language technology (101195233): Granted funding: 334,233 EUR, source: Research Council of Finland, Role: Scientific Investigator
Feb 2025 – Sep 2026	Grant name: High Performance Language Technologies (101070350) Granted funding: 3,880,687 EUR, source: HORIZON Europe Role: Scientific Investigator
Sep 2023 – now	Grant name: Edustories: Using artificial intelligence and virtual reality in developing competencies to solve challenging behavior in the classroom (TQ01000030) Granted funding: 420,000 EUR, source: Technology Agency of the Czech Republic Role: Scientific Investigator Note: key role in interdisciplinary collaboration — driving the integration of computer science technologies to another research discipline (Education)
Sep 2022 – Feb 2024	Grant name: Intelligent Back-Office (CZ.01.1.02/0.0/0.0/21_374/0026711): Granted funding: 180,000 EUR, source: Ministry of Industry and Trade of Czech Republic, Role: Research Technical Lead Note: key role in industrial innovation — designing and coordinating the translation of research into a commercialised product

Prizes and Awards

Aug. 2026	Awarded a Runner-Up recognition for the Informatics Europe Best Dissertation Award
July 2026	Spotlight oral presentation at ICML Mechanistic Interpretability workshop (~2% of submissions)
Oct. 2025	Received Prorector's Award for Outstanding PhD Research
May 2025	Received Dean's PhD Award for the defended Dissertation thesis (~5% of defended)
Feb. 2025	Awarded Summa Cum Laude for the defended Dissertation thesis (2–5% of defended)
Nov. 2024	Received Outstanding Reviewer Award at EMNLP 2024
Apr. 2024	Named Outstanding PhD Researcher by the Research Board of the Faculty of Informatics, MU
July 2023	First author of Best paper at ACL 2023 NLRSE (among 12 accepted)
May 2023	First author of Best paper at EACL 2023 SlavicNLP (among 26 accepted)
July 2022	First co-author of a winning submission in NAACL DADC: Better Training Data challenge
Nov. 2021	First author of Best-performing metric at WMT 2021 Metrics shared task (among 18 submissions)

References

1. **Pontus Stenetorp**, Deputy Directory of the Centre for AI and Full Professor at University College London
email: p.stenetorp@cs.ucl.ac.uk
2. **Jörg Tiedemann**, Full Professor and former Principal Investigator at University of Helsinki
email: jorg.tiedemann@helsinki.fi
3. **Petr Sojka**, Associate Professor and PhD advisor, Faculty of informatics at Masaryk University,
email: sojka@fi.muni.cz