István Üveges, PhD

Computational Linguist & Machine Learning Researcher

info@uvegesai.com www.uvegesai.com /in/istvan-uveges ORCID: 0000-0001-5897-9379

Profile

Computational linguist and ML researcher (PhD) focusing on Natural Language Processing, from classical models to large language models, with a strong emphasis on clarity, interpretability, and practical applications. I design and deploy end-to-end NLP pipelines, build domain-specific models (legal, political, financial), and work on explainability and robustness to ensure trustworthy outcomes. Beyond research, I lead outreach efforts to make AI and NLP accessible to wider audiences.

Current Role

HUN-REN Centre for Social Sciences (POLTEXTLAB) — NLP & ML (current)
Applied NLP for legal/political texts (classification, sentiment/emotion, plain language). Model transparency with SHAP-based analyses and robustness checks across domains and noise types. End-to-end pipelines from data curation to serving/demo.

Education

PhD in Linguistics — University of Szeged (2024)
Dissertation: *Közérthetőség és automatizáció* (*Comprehensibility and Automation: Plain Language in the Era of Digitalization*). DOI

BSc in Computer Science — University of Szeged (2024)

Skills (selected)

NLP/ML: Transformers (BERT, XLM-R), sentiment & emotion analysis, plain-language detection, explainability (SHAP), augmentation (EDA, back-translation, LLM), clustering & topics (UMAP/t-SNE, HDBSCAN, BERTopic).
Stack: Python, PyTorch, Hugging Face, scikit-learn, spaCy, FastAPI, PostgreSQL, Docker.
MLOps/Eval: Optuna/Bayesian search, LoRA/PEFT, mixed precision, model cards, latency/throughput, experiment tracking.

Python PyTorch Hugging Face spaCy FastAPI Docker PostgreSQL SHAP UMAP HDBSCAN

Selected Publications

- Üveges, I., Ring, O. (2025). *Evaluating the Impact of Synthetic Data on Emotion Classification: A Linguistic and Structural Analysis*. Information. DOI
- Ring, O., Szabó, M.K., Guba, Cs., Váradi, B., Üveges, I. (2024). *Approaches to Sentiment Analysis of Hungarian Political News at the Sentence Level*. Language Resources and Evaluation. DOI
- Csányi, G.M., Lakatos, D., Üveges, I., et al. (2024). *From Fact Drafts to Operational Systems: Semantic Search in Legal Decisions Using Fact Drafts*. BDCC. DOI

- Vági, R., Üveges, I., et al. (2024). *Increasing Access to Legal Information with Unsupervised Solutions*. Hungarian Journal of Legal Studies. DOI
- Vági, R., Üveges, I. (2024). *Laws Clearly: Large Language Models and Plain Language Transformation*. Magyar Nyelvőr. DOI
- Üveges, I., Ring, O. (2023). *HunEmBERT: A Fine-Tuned BERT-Model for Classifying Sentiment and Emotion in Political Communication*. IEEE Access. DOI

Full list: Google Scholar | uvegesai.com/publications

Outreach (selected)

Lead author of the *Tech & AI* column at *Constitutional Discourse*, writing accessible articles on AI, NLP and technology policy for a broad audience. Link