Jordy Van Landeghem

Document AI Researcher

✉ Fazantenweg 31
2560 Nijlen (Belgium)
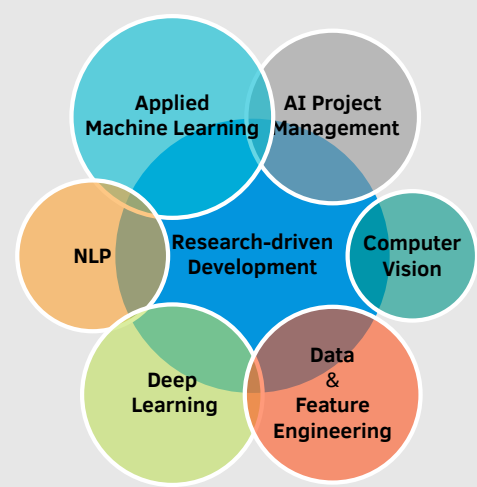
✉ jordy.vlan@gmail.com

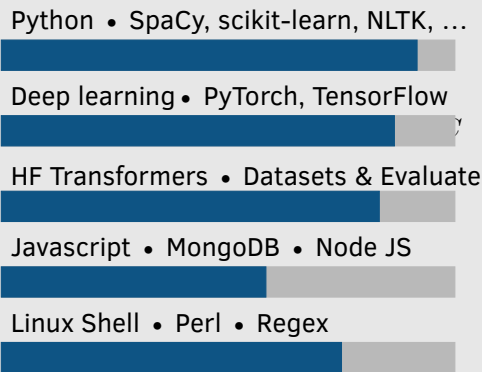📱 (+32) 489642832

in jordy-van-landeghem-3b1166b3

⬡ Jordy-VL

Skills

Applied Machine Learning · **AI Project Management** · **NLP** · **Research-driven Development** · **Computer Vision** · **Deep Learning** · **Data & Feature Engineering**

Programming

Python • SpaCy, scikit-learn, NLTK, …

Deep learning • PyTorch, TensorFlow

HF Transformers • Datasets & Evaluate

Javascript • MongoDB • Node JS

Linux Shell • Perl • Regex

Languages

Dutch	● ● ● ● ●
English	● ● ● ● ●
Spanish	● ● ● ● ○
French	● ● ● ● ○
German	● ● ● ○ ○
Portuguese	● ● ● ○ ○

Experience

2024-now **Senior Software Engineer, Machine Learning** **Instabase [Remote]**

- Developed *custom LLMs*, training and benchmarking frameworks.
- Integrated *LLM/VLM RAG* workflows to enhance product capabilities.
- Engineered *Agentic Search* for large-scale (1M+) document repositories.
- Spearheaded model development initiatives as the technical lead.

2017-2024 **Lead AI Research Engineer** **Contract.fit [Brussels]**

- *AI Engineering*: Document Understanding (NLP & CV)
- *ML Operations*: Design and implementation of production-grade pipelines.
- *Mentorship*: Supervised 11 Master's AI theses.

2017 **Research Intern**
Language Modelling Algorithms **Nuance Communications [Aachen]**

- Performed research into advanced regularisation techniques for RNN LMs.
- Implemented biLSTM induction of character-based word embeddings.

2016-2017 **Research Intern**
NLP Development for Dialogue Systems **Oracle [Barcelona]**

- Investigated Seq2Seq NNs for Chatbot & Virtual Assistant Technology.

Education

2020 - 2024 **Ph.D in Engineering Sciences, Computer Science** **KU Leuven**
PhD Thesis: *Intelligent Automation for AI-driven Document Understanding*
TA: Data Structures & Algorithms, Popular Science Communication.

2015 - 2017 **Master of Science, Artificial Intelligence:**
Speech and Language Technology (Magna Cum Laude) **KU Leuven**
Thesis: *Sequence-to-Sequence Learning for End-to-End Dialogue Systems*

2014 - 2015 **Master of Arts, Linguistics:**
Corpus- and usage-based Methodologies (Magna cum Laude) **KU Leuven**
Thesis: *-ing seen through its Dutch Translation Equivalents: a contrastive Parallel Treebank based study*

Research Publications

Zhou, Han, Van Landeghem Jordy et al.*A Novel Characterization of the Population Area Under the Risk Coverage Curve (AURC) and Rates of Finite Sample Estimators*. ICML 2025.

Van Landeghem, Jordy, et al. *DistilDoc: Knowledge Distillation for Visually-Rich Document Applications*. ICDAR 2024.

Van Landeghem, Jordy, Sanket Biswas, Matthew Blaschko, and Marie-Francine Moens. *Beyond Document Page Classification: Design, Datasets, and Challenges*. Winter Conference on Applications of Computer Vision (WACV), 2023.

Van Landeghem, Jordy, et al. *Document Understanding Dataset and Evaluation (DUDE)*. In International Conference for Computer Vision (ICCV), 2023.

Van Landeghem, Jordy, et al. *ICDAR 2023 Competition on Document UnderstanDing of Everything (DUDE)*. In ICDAR 2023.

Van Landeghem, Jordy, Matthew Blaschko, Bertrand Anckaert, and Marie-Francine Moens. *Benchmarking Scalable Predictive Uncertainty in Text Classification*. IEEE Access, 2022.

Van Landeghem, Jordy, Matthew Blaschko, Bertrand Anckaert, and Marie-Francine Moens. *Predictive Uncertainty for Probabilistic Novelty Detection in Text Classification.* In Proceedings ICML 2020 Workshop on Uncertainty and Robustness in Deep Learning (ICML), 2020.

Francis, Sumam, Jordy Van Landeghem, and Marie-Francine Moens, *Transfer Learning for Named Entity Recognition in Financial & Biomedical Documents*. Information 10.8, 2019.