Jordy Van Landeghem

Document AI Researcher

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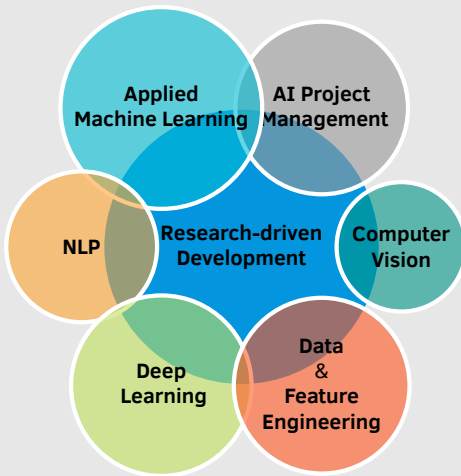
✉ jordy.vlan@gmail.com

☎ (+32) 489642832

in jordy-van-landeghem-3b1166b3

○ Jordy-VL

Skills

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

Jun 2017 - now

AI/NLP Research Engineer   Contract.fit [Brussels]

- Researching and prototyping the state-of-the-art in NLP and Computer Vision to advance Document Understanding.
- Designed and implemented production-grade Machine Learning pipelines for Document Classification and Information Extraction.
- Applied end-to-end AI engineering in application domains like insurance, finance, legal & creating building blocks of a SaaS product.
- Supervised 11 CS/AI Master's thesis internships

Mar 2017 - June 2017

Research Intern
Language Modelling Algorithms   Nuance Communications [Aachen]

- Performed research into advanced regularisation techniques for RNN-based Language Models.
- Contributed to an in-house TensorFlow toolkit and implemented biLSTM induction of character-based word embeddings.

Aug 2016 - Feb 2017

Research Intern
NLP Development for Dialogue Systems   Oracle [Barcelona]

- Investigated Sequence-to-Sequence Neural Networks for Dialogue Modelling, Chatbot & Virtual Assistant Technology.
- Presented results at *the 27th Conference of Computational Linguistics in the Netherlands (CLIN27)*.

Education

2020 - 03/24 **Ph.D in Engineering Sciences, Computer Science**   KU Leuven
Project: ***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

Van Landeghem, Jordy, et al. *DistilDoc: Knowledge Distillation for Visually-Rich Document Applications*. CVPR 2024 (under review).

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 17th International Conference on Document Analysis and Recognition, 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.