Jeroen Van Goey

Machine Learning for Biology

Cape Town, South Africa
+27 72 458 2796
jeroen@vangoey.be
linkedin.com/in/jeroenvangoey
jeroen.vangoey.be/

Work experience

Staff Research Engineer in BioAI
2022-08 - Present

InstaDeep

- Team lead and hiring manager for BioAI research projects, leading two cross-disciplinary research teams:
- ***De novo* peptide sequencing** project in collaboration with the Technical University of Denmark, leveraging deep learning to innovate peptide discovery and analysis.
- **Signal peptides research** in collaboration with BioNTech, focusing on improving secretion efficiency crucial to immunotherapy and vaccine development using generative AI.
- Responsible for recruitment and team expansion, overseeing BioAI position hires for offices in Cape Town and Kigali, ensuring talent alignment with cutting-edge research initiatives and company goals.
- Collaborating with academic and industry partners to drive innovation, scientific rigor, and project success.

Senior Software Development Engineer - Machine Learning
2020-02 - 2022-08

Barco

- Established and optimized a machine learning production pipeline, leveraging Tensorflow Extended, to train deep learning models on multispectral images for the detection and classification of skin cancer melanomas.
- Collaborated closely with the research team and development teams responsible for the cloud-based backend, as well as the mobile and web frontends, ensuring seamless integration and efficient workflow.

Bioinformatics Researcher & Python + R Platform manager
2018-08 - 2020-01

BASF Agricultural Solutions (before divestment: Bayer CropScience)

- Enhanced and expanded the Python/Perl based data analysis platform at BASF Agricultural Solutions, significantly boosting data analysis capabilities and improving efficiency.
- Provided technical and functional support to developers and data scientists within the BASF network, ensuring smooth operation and optimal utilization of the platform.
- Conducted training sessions on Python basics, NumPy, Pandas, and BioPython for data scientists, enabling them to leverage these tools effectively in their work.
- Assumed additional responsibilities in July 2019 as the administrator of the internal R-Platform and the Atlassian stack (BitBucket, Jira, Bamboo, etc.), ensuring their smooth functioning and providing technical support to users.

Bioinformatics Software Developer
2011-09 - 2018-01

BioMérieux (before acquisition: Applied Maths)

- Developed custom scripts and extensions in Python for clients in the clinical, academic, and government sectors, enhancing the functionality and usability of the BioNumerics bioinformatics software suite.
- Collaborated with cross-functional teams to understand client requirements and translate them into efficient and effective software solutions.
- Created the Functional Genotyping Plugin and the Whole Genome Multi Locus Sequence Typing Plugin, enabling clients to analyze biological data with greater accuracy and efficiency.
- Conducted thorough testing and debugging of software, ensuring the reliability and accuracy of results generated by the plugins.

Java Programmer
2009-02 - 2009-03

Biogazelle

- Interned at Biogazelle, where I wrote a Java extension on the Eclipse BIRT platform for real-time qPCR analytics software.

Education

Microdegree Machine Learning and Deep Learning

2018 - 2019

Hogeschool West-Vlaanderen - Vlaamse Autonome Hogeschool

Credit program where we learned about Convolutional Neural Networks, Autoencoders, Generative Adverserial Networks, Recommendation systems, Long Short-term memory networks and Reinforcement Learning.

M.Sc. Bioinformatics

2010 - 2012

Katholieke Universiteit Leuven

Partial credits obtained in Linear Algebra, Neural Networks, Support Vector Machines, ...

M.Sc. Biology

1997 - 2004

Universiteit Antwerpen

Thesis: Comparison of methods to prioritize candidate disease genes according to their presumed involvement in human hereditary diseases.

Skills

Python

▰▰▰▰▰▰▰▰▰▱

10+ years of professional programming experience.
Knowledge of Jax, Pytorch, NumPy, Pandas,
Open source contributions to CPython, BioPython,...
'Expert' assesment by PluralsightIQ.
Finalist in the Cisco Python Challenge 2018.

Bioinformatics

▰▰▰▰▰▰▰▱▱▱

6 years of professional experience developing bioinformatics software. Managing both open source and internally developed bioinformatics tools on a HPC infrastructure. Knowledge of BLAST, ClustalW, Snakemake, BioNumerics,...

Deep Learning

▰▰▰▰▰▰▱▱▱▱

ML lead on InstaNovo: accurate, database-free peptide identification for large scale proteomics experiments
Brought a machine learning pipeline for skin cancer detection from research phase to production.
Created proof-of-concept setup of a reproducible machine learning pipeline using Kubernetes / Kubeflow Pipelines.

Publications

De novo peptide sequencing with InstaNovo: Accurate, database-free peptide identification for large scale proteomics experiments

biorXiv

Kevin Eloff, Konstantinos Kalogeropoulos, Oliver Morell, Amandla Mabona, Jakob Berg Jespersen, Wesley Williams, Sam P. B. van Beljouw, Marcin Skwark, Andreas Hougaard Laustsen, Stan J. J. Brouns, Anne Ljungars, Erwin M. Schoof, Jeroen Van Goey, Ulrich auf dem Keller, Karim Beguir, Nicolas Lopez Carranza, Timothy P. Jenkins

afkSNP: Assembly-free K-mer based SNP comparison of bacterial WGS samples

BeNeLux Bioinformatics Conference

Jeroen Van Goey, Hannes Pousseele, Philip Supply, Stefan Niemann