

# Jaco du Toit

Cape Town, South Africa | +27 78 834 4936

jacowdutoit11@gmail.com

 LinkedIn |  Github |  Google Scholar

## Summary

I am a computer scientist and researcher with a strong background in machine learning, natural language processing (NLP), and large language models. My academic background and professional experience have enabled me to work on a range of projects, from building NLP systems to developing machine learning pipelines for real-world applications. I am excited by the rapid advancements in AI systems, but concerned about the risks they may pose. Therefore, I am dedicated to exploring AI safety, interpretability, and alignment in my future research to ensure that AI benefits all humanity.

## Education

### Stellenbosch University

Jan 2023 - March 2024

MSc Computer Science

- **Thesis:** Hierarchical Text Classification with Transformer-based Language Models.

Cum Laude (85%)

### Stellenbosch University

Jan 2022 - Dec 2022

BSc (Hons) Computer Science

- Focus on machine learning with mini thesis in machine learning and natural language processing.

Cum Laude (86%)

### Stellenbosch University

Jan 2019 - Dec 2021

BCom Mathematical Sciences (Data Science)

- Mathematics and Computer Science.

Cum Laude (87%)

## Experience

### AI Safety Engineering Taskforce

Jan 2025 - Present

Research Engineer.

- Developing safety evaluations for large language models.

### Matogen AI

May 2024 - Present

Data Scientist.

- Developing machine learning applications in finance, retail, and agriculture industries.

### Stellenbosch University

Jan 2022 - July 2022

Computer Science Teaching Assistant.

## Publications

- du Toit, J. and Dunaiski, M. (2024). Prompt Tuning Discriminative Language Models for Hierarchical Text Classification. Natural Language Processing. <https://doi.org/10.1017/nlp.2024.51>
- du Toit, J., Redelinghuys, H., and Dunaiski, M. (2024). Introducing Three New Benchmark Datasets for Hierarchical Text Classification. arXiv preprint arXiv:2411.19119. <https://arxiv.org/abs/2411.19119>
- du Toit, J., Dunaiski, M. (2023). Hierarchical Text Classification Using Language Models with Global Label-Wise Attention Mechanisms. In: Artificial Intelligence Research. SACAIR 2023. Communications in Computer and Information Science, vol 1976. Springer, Cham. [https://doi.org/10.1007/978-3-031-49002-6\\_18](https://doi.org/10.1007/978-3-031-49002-6_18)

## Skills

**PROGRAMMING LANGUAGES** *Experienced:* Python *Familiar:* C | Java | R | Javascript | SQL

**SOFTWARE DEVELOPMENT** GIT | Docker | CLI | Agile Methodology

**FRAMEWORKS & LIBRARIES** PyTorch | TensorFlow | HuggingFace | Pandas | PySpark | Scikit-learn

**LANGUAGES** *Fluent:* English | Afrikaans *Beginner:* Dutch | German | Spanish

## Awards, Honors, and Scholarships

---

<b>School for Data Science and Computational Thinking</b> Scholarship for master's studies.	2023 - 2024
<b>Stellenbosch University</b> Faculty of Science Postgraduate Research Conference: Best Presentation.	2023
<b>Stellenbosch University</b> Rector's award for academic excellence. Awarded for being in top 10 of faculty.	2022
<b>Stellenbosch University</b> Academic merit award. Awarded for being in top 5% of faculty.	2019 - 2021
<b>Golden Key International Honour Society</b> Awarded for being in top 15% of class.	2019

## Additional Education

---

<b>Condor Initiative</b> AI Safety Camp Cape Town • One week upskilling course in AI safety.	November 2024
<b>BlueDot Impact</b> AI Safety Fundamentals Alignment Course • 3-month part-time course covering fundamentals of AI safety with a research project in final month.	October 2024 - Present
<b>Udemy</b> Artificial Intelligence A-Z • Online Machine Learning course through Udemy.	Jan 2022 - March 2022
<b>Stanford Online</b> Machine Learning • Online Machine Learning course from Stanford University through Coursera.	Jan 2021 - March 2021

## Extracurricular Activities

---

<b>AI Safety Cape Town</b> Weekly reading group on AI safety topics.	June 2024 - Present
<b>Stellenbosch University</b> Co-supervisor for Computer Science student research project.	July 2023 - Dec 2023
<b>Language and Speech lab, Stellenbosch University</b> Weekly reading group on deep learning for language and speech.	Jan 2023 - Dec 2023
<b>Eco Maties, Stellenbosch University</b> Executive member for the environmental society of Stellenbosch University.	Jan 2020 - Dec 2020
<b>Dagbreek Residence, Stellenbosch University</b> Academic mentor for 6 first-year Computer Science students.	Jan 2020 - Dec 2020
<b>Paul Roos Gymnasium</b> Mentor for 3 incoming high-school students.	Jan 2018 - Dec 2018