🌐 **JANSOBUS.COM**

## CONTACTS

✉ sobus.jan@gmail.com

📞 +61 434 848 648

📍 Burpengary East,
4505 QLD, Australia

in Jan Sobus

## SKILLS

- **Generative AI:**
  - LangChain
  - Vector and Graph DBs, RAG
  - Agentic systems (DSPy)
  - Local models with Ollama
  - OpenAI API integration
- **Deep learning:**
  - PyTorch
  - Tensorflow + Keras
  - Hugging Face
  - Fast.ai
- **APIs, messaging and UI:**
  - FastAPI
  - Flask
  - Streamlit
  - gRPC
  - MQTT/AMQP
- **Containerization and CI/CD:**
  - Docker
  - Azure Pipelines
  - GitHub actions
  - Kubernetes
- **Cloud:**
  - Azure
  - AWS
  - Databricks
- **Data analysis and ML:**
  - Pandas/Numpy
  - Scikit-learn
  - XGBoost
  - SciPy
- **Visualization:**
  - Plotly
  - Matplotlib
  - Leaflet
  - PowerBI
- **Project management:**
  - Git
  - Jira
  - Azure DevOps
- **Storage**
  - Postgres
  - Redis
- **Other languages:**
  - C
  - R
  - MATLAB
  - HTML + CSS

# JAN SOBUS, PHD

## ML ENGINEER, ARCHITECT, RESEARCHER

Experienced architect with research background and wide scope of experience that spans domains of machine learning, software engineering, data science and physics. I combine strong theoretical skillset with developer experience in production environments. My goal is to provide tailored data-science and AI solutions addressing real life challenges.

## RECENT EXPERIENCE

### DATA SCIENCE ARCHITECT
*September 2023 – Now*
**CATERPILLAR OF AUSTRALIA** – BRISBANE, AUSTRALIA
In the current architect role I am designing deployment patterns, architecture and coding standards for ML components to be developed by multiple Data Science teams across the enterprise. I am also helping with exploration of new areas of artificial intelligence.

- Currently supervising several independent Data Science project spanning time series analysis, image processing, LLMs and reinforcement learning.
- Designed and built a mine digital twin platform that enables quick prototyping of data science models, while using same interfaces to communicate with field/simulators as the production infrastructure – FastAPI + Docker + Redis + PostGIS

### DATA SCIENCE TEAM LEAD
*December 2021 – September 2023*
**CATERPILLAR OF AUSTRALIA** – BRISBANE, AUSTRALIA
Led a team of five Data Scientists that provided analytical and machine learning solutions for the software Caterpillar software and product teams

- Designed and built a novel assignment engine for optimizing mine operations by using ensemble of ML models form different families (supervised learning, unsupervised learning, reinforcement learning) opening way towards use of electric fleets in Caterpillar products.
- Performed the consolidation and migration of Data Science codebase into Azure Repos and developed automated Azure build pipelines that reduced delivery time to engineering teams by 70%.

### MACHINE LEARNING SOFTWARE ENGINEER
*July 2020 – October 2021*
**EXCI.AI** – MAROOCHYDORE, AUSTRALIA
At EXCI.AI I was part of the team that has built a system for early fire detection based on ground and satellite data. I was responsible for machine learning component of the system, which has reached capacity to detect fires up to 10 km from the camera in less than 3 minutes and pinpoint them on the map.

- Built a multi-model machine learning pipeline for smoke detection in near-real time.
- Successfully integrated gigabytes of data from land cameras, satellites and external data streams.
- Implanted the complete ML solution into the AWS hosted production system based on custom-built microservices connected with message queues and RESTful APIs.

## EDUCATION

### PHD IN PHYSICS
*2013 – 2016*
**ADAM MICKIEWICZ UNIVERSITY (POLAND) & EPFL (SWITZERLAND)**
My PhD project was focused on development of novel hybrid organic-inorganic solar cells (Dye Sensitized Solar Cells) as a potential replacement for widespread silicon photovoltaics. It covered both experimental (cell manufacturing and measurements) and simulation (modelling photophysical and optoelectronic processes) aspects of those devices. I continued my academic career as a postdoc at the University of Queensland until 2020, publishing over 15 papers (up to Nature group journals), collaborating in multiple international projects and promoting two PhD students.