Tomasz Mycielski

linkedin.com/in/tmycielski | github.com/mycielski | tom@myciel.ski | +48 606 273 309

Summary

Versatile engineer with experience building end-to-end data solutions, from ETL pipelines and infrastructure-as-code, through analysis, to visualizing and reporting actionable insights. Seeking to leverage a broad set of skills to solve interesting problems across diverse industries.

Education

Warsaw University of Technology Warsaw, Poland

Master of Science in Computer Science *2025*

GPA 4.24 (scale 2 → 5)
Scholarship for Academic Performance

Experience

Data Scientist III 2025 – Present

Zest AI *Burbank, California*

- Forecasting revenue from credit lending across credit card, vehicle, mortgage, and personal loans
- Designing and implementing metrics for evaluating models' predictive power
- Visualizing key changes over time in populations of borrowers to facilitate identifying significant trends

Data Engineer VII 2024 – X 2024

Taiwan Semiconductor Manufacturing Company *Hsinchu, Taiwan*

- Monitored Overall Equipment Effectiveness in a 300 mm wafer fab
- Implemented ETL workflows for processing 20TB of logs daily with Dbt, Trino, Iceberg
- Orchestrated data pipelines with Airflow
- Containerized and deployed data-wrangling code to Kubernetes

Software Engineer V 2022 – VI 2024

Flobotics Consulting Inc. *Munster, Indiana*

- Provided solutions to clients having thoroughly understood their business needs
- Deployed to cloud (mostly AWS) and on-premise environments, often with Terraform
- Worked with both small and large companies – including DHL, TransPerfect, PathGroup, Docplanner
- Kept doing freelance/contract work for Flobotics to this day

Security Operations Center Analyst VII 2021 – X 2021

T-Mobile *Warsaw, Poland*

- Developed an event-driven system for automated malware detonation
- Responded to security incidents
- Analyzed malicious samples for threat intelligence
- Collaborated on reports and presentations for management

Web Developer 2016 – 2022

Freelance *Europe*

- Built websites for small businesses, entrepreneurs, people wanting to increase their on-line footprint
- Handled the business end-to-end, from marketing my services, to designing sites, through implementation, to continued support

Skills and Tools

Technical: Python, SQL, Bash, Pandas, NumPy, Matplotlib, Playwright, Selenium, Burp Suite, AWS, Terraform, Linux, Docker, Git, Kubernetes, RabbitMQ
Language: Polish (Native), English (Fluent), French (Basic)
Laboratory: Machine Learning, Signal and Image Processing (DSP), Exploratory Data Analysis (EDA), Regular Expressions (regex), Web Scraping, Open-Source Intelligence (OSINT), Cybersecurity
Interests: Piano, Golf, Climbing

Selected Projects

Digitization of Apartment Floor Plans | *OpenCV, scikit-image, PyTorch*

- Implemented computer vision algorithms to parse images of floor plans
- YOLO for detecting features (doors, windows, balconies, etc)
- Mask R-CNN for segmenting plans into constituent rooms

Invalid Bill of Lading Detection | *scikit-image, OpenCV*

- Detected key regions on scanned BOLs
- Validated contents of regions using traditional and neural network image processing techniques in an ensemble model
- Assigned probability of document being invalid to each BOL

Video Content Automation Pipeline for Lead Generation | *ffmpeg, AWS (EC2, S3, SQS), Hubspot, OpenAI*

- Engineered a serverless pipeline for creating and sending personalized outreach videos to leads
- Integrated Azure OpenAI for NLP-driven script generation and Playwright for automated screen recording
- Utilized ffmpeg on AWS EC2 for cost-effective, at-scale video processing and editing

Recurrent Neural Network from scratch in Julia | *Julia, Jupyter Notebooks*

- Implemented an RNN from scratch, including my own automatic differentiation (BPTT) library
- Achieved over 96% accuracy on the MNIST test dataset

Insurance claim optimization from Electronic Health Records | *GCP (Firestore, Compute Engine), Selenium*

- Created a system for automated extraction of EHRs from management systems
- Documented the system's API to facilitate creation of new connector plugins
- Stored gathered data in GCP Firestore for use by machine learning models
- Anonymized personally identifiable information for processing
- Containerized scrapers and deployed to GCP Compute Engine

Fuel theft detection for lorry fleet operator | *REST API, Microsoft SQL Server, PowerBI*

- Developed an ETL solution aggregating fuel data from multiple sources into an SQL database
- Implemented anomaly detection algorithms to identify potential fuel theft from sensor readings
- Cleaned, preprocessed, and visualized data for analysis and reporting

Medical diagnosis document understanding | *AWS (Textract, S3, SQS, Lambda, SNS)*

- Built a serverless data processing solution to digitize up to 100 000 pages of medical documents monthly
- Leveraged AWS Textract to parse PDFs and extract relevant diagnosis details
- Used AWS Lambda, SQS, and S3 to ensure scalability and reliability

Market intelligence using WhatsApp | *WhatsApp Cloud API, AWS (Lambda, S3, SQS)*

- Designed system for en masse polling of medical practices to check what system they use for patient booking
- Implemented a bot using the WhatsApp Cloud API and AWS Lambda to automate data collection
- Analyzed aggregated data with Pandas to deliver an actionable, data-backed report to the client

Land and mortgage register scraper | *Playwright, httpx, Docker, k8s, RabbitMQ, MinIO*

- Built a web scraping solution to gather judicial data using Playwright and httpx
- Designed and deployed a distributed, partition-tolerant system to circumvent web application firewalls and rate limiting
- Managed deployment across multiple cloud providers and an on-premise cluster

Marketing and Sales Helper Chatbot | *Azure OpenAI*

- Built a knowledge base from internal pitch decks, case studies, marketing materials
- Created retrieval augmented generation (RAG) on these documents
- Deployed an Azure OpenAI GPT as a means to interface with the knowledge base for non-technical sales and marketing staff, and get help when preparing promotional materials and during sales calls

Mode S protocol extension for message authentication | *SciPy, HackRF, RTL-SDR*

- Designed an extension to Mode S for addressing lack of message authentication
- The extension is fully backwards compatible with existing specification of the protocols Mode A/B/S
- Developed a proof-of-concept implementation in Python for use with off-the-shelf consumer SDR hardware