

Szymon Socha

<https://szymonsocha.github.io/>

✉ [upon request](#) | 📱 [upon request](#)
in [szymon-socha](#) | 🌐 [szymonsocha](#)

EXPERIENCE

Data Scientist | *Python, PySpark, Machine Learning, AML*

Feb 2024 – present

PwC Czech Republic / TEAM International

Remote, PL

Data-driven support for the banking industry in the field of **Anti-Money Laundering** (AML) and financial crime

- Developing and enhancing Anti-Money Laundering (**AML**) system code using Python and **PySpark**
- Optimizing **transaction monitoring** (TM) scenarios by applying statistical methods and machine learning algorithms
- Member of the Financial Crime Technology team

Data Scientist | *Python, Machine Learning, NLP, Time Series Forecasting, LLM, Research*

Mar 2023 – Dec 2023

WeSub

Warsaw (Remote), PL

A member of the team conducting an R&D project aimed at developing innovative models for product categorization and value prediction using machine learning and artificial intelligence algorithms

- Developing price prediction models (**Time Series Forecasting**) by leveraging Natural Language Processing (NLP) and **transfer learning** features
- Researching cutting-edge techniques for employing and **fine-tuning of Large Language Models** (LLMs)
- Creating machine learning models explaining economic phenomena
- The National Centre for Research and Development (NCBR) project

Data Scientist (Junior) | *R, SQL, BigQuery, ETL*

Jan 2022 – Apr 2023

Flying Wild Hog

Warsaw (Remote), PL

Analysing telemetry data from video games with a focus on improving User Experience (UX)

- Analyzing video games telemetry data using **R** and **SQL**
- Developing dashboards and reports in **RMarkdown** and **Shiny**
- Developing **ETL** processes with **BigQuery** and **MongoDB**
- Collaborating with UX Research team members in combining quantitative and qualitative research findings

Data Science Assistant | *Python, R, Econometric Modelling*

Sep 2021 – Nov 2021

Publicis Groupe

Warsaw, PL

Developing tool for econometric modeling project

- Data cleaning and processing using **Python** and **R** and building custom R tools for econometric analysis

EDUCATION

Master of Science in Data Science and Business Analytics

University of Warsaw

GPA: 5/5, **Awards:** Dean's List (Top 7%)

Thesis: *Should You Care about Tools in Fake News Detection? A Comparative Study of Cutting-Edge Embedding Methods and Classification Algorithms*

Courses: Machine Learning I & II, Unsupervised Learning, Advanced Econometrics, Applied Finance, Statistics, Python, Advanced Programming in R, Big Data Analytics, Algorithms for Data Science, Bayesian Time Series Forecasting

Bachelor of Science in Computer Science and Econometrics

University of Warsaw

Thesis: *The Impact of the COVID-19 Pandemic on Music Streaming Consumption*

Courses: Machine Learning, Econometrics, Linear Algebra, Statistics, Time Series, Probability Calculus, Finance, Python, R, SQL, C++

SELECTED PROJECTS

Option pricing Rcpp package | *C++, R, Rcpp, Rmd*

- Path-dependent European style up-and-out call option pricing with Monte Carlo and Rcpp package

TECHNICAL SKILLS

Languages: Python, R, SQL (BigQuery, KQL)

Tools: Git (GitLab, Bitbucket), FastAPI, Google Cloud Platform, PyCharm, RStudio, Ubuntu

Libraries: PyTorch, PyTorch Lightning, HuggingFace, PySpark, sklearn, pandas, NumPy, tidyverse, dplyr, ggplot2

Expertise: machine learning, natural language processing (NLP), time series forecasting, econometric data modelling, clusterization, dimension reduction, statistics, hypothesis testing