

# Szymon Socha

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

✉ [szymon.socha13@gmail.com](mailto:szymon.socha13@gmail.com) | ☎ +48-603-820-390

🌐 [szymon-socha](#) | 🐙 [szymonsocha](#)

## EDUCATION

University of Warsaw

Warsaw, PL

*Master of Science in Data Science and Business Analytics*

2021 – 2023

**GPA:** 5/5, **Awards:** Dean's List 2022 (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, Machine Learning II, Unsupervised Learning, Advanced Econometrics, Applied Finance, Statistics, Python, Advanced Programming in R, Big Data Analytics, Algorithms for Data Science, Bayesian Time Series Forecasting

University of Warsaw

Warsaw, PL

*Bachelor of Science in Computer Science and Econometrics*

2018 – 2021

**Thesis:** Investigating the impact of the COVID-19 pandemic on music streaming consumption, utilizing Google Trends activity changes and Spotify statistics. An econometric model is constructed and evaluated using R and Python

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

## EXPERIENCE

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

Mar 2023 – Present

*WeSub*

*Warsaw (Remote), PL*

- Conducting R&D to develop new models for product categorization and value prediction within a circular economy framework, using machine learning and artificial intelligence algorithms
- Developing a range of **Time Series Forecasting** models 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 **NLP** models for text generation and classification
- 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*

- 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
- Working closely with different game development teams and editorial teams around the world in order to understand their requirements and questions related to a project

**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

## 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, 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

I agree to the processing of personal data provided in this document for realising the recruitment process pursuant to the Personal Data Protection Act of 10 May 2018 (Journal of Laws 2018, item 1000) and in agreement with Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).