**WriteUp**

Hi, I am Marcin, a 24-year-old from Katowice, Poland. From a young age, I was a math nerd. However, I was always frustrated by the theory-oriented nature of my beloved field. I wanted to find out what maths can teach me about the real world. As a result, I have ended up majoring in Informatics and specializing in Machine Learning (ML). My academic career taught me how to channel my math skills to derive valuable insights through data analysis. Right now, I cannot wait to see how can I use my data science toolkit to benefit your company.

**Interests**

* Data Science
* Artificial Intelligence
* Machine Learning
* Deep Learning
* Natural Language Processing
* Software Engineering

**Education**

**The University of Edinburgh**

I have completed a five-year undergraduate Integrated Master's program worth 300 ECTS credits at the University of Edinburgh. My alma mater is among the top 15 world universities at the time of my graduation, according to the [QS ranking](https://www.topuniversities.com/university-rankings/world-university-rankings/2023).

**International Baccalaureate Diploma**

In high school, I challenged myself by rejecting the standard Polish Matura and joining the alternative IB programme. The choice was difficult as I am a Polish native, while IB offers teaching fully in English. In the end, I completed the programme with a maximum possible grade of 7 in 4 of my subjects (Mathematics, Physics, Business and Management, Polish Literature), and grade 6 in the remaining 2 (English and Chemistry).

In my prior education, I have always been a high performer. In my secondary school, I was a laureate of three distinct Regional Subject Olympiads in the Silesian Voivodship, namely in Mathematics, Physics and History.

**Skills**

**Data Science**

Python with its scientific programming toolkit is my favourite programming language. I use scientific Python whether it concerns directly extracting statistical summaries from data, or pre-processing it for downstream machine learning tasks. I have worked with Computer Vision, Natural Language and Time-Series datasets. On occasions, my friends asked me for help when their projects involved data analytic tasks in fields as distinct as Physics and Linguistics. Generally, you name a field with data, and I got your results. Concerning technical quirks, I naturally look for ways to vectorize Python code to get performance speed-ups. In the field of data science, I especially love data visualisation techniques.

### **Deep Learning**

In my studies, using NumPy, I have implemented from scratch (forward and back propagation) most canonical components of modern neural networks from simple multi-layer perceptrons to complex convolutional layers, recurrent networks, batch normalization, residual connections, etc. Due to the above first principles work, I have strong background concerning the deep learning concepts and the inner workings of higher-level libraries such as PyTorch.

Reading Andrej Karpathy's [The Unreasonable Effectiveness of Recurrent Neural Networks](https://karpathy.github.io/2015/05/21/rnn-effectiveness/) blog post, inspired me to write my own modern PyTorch version of the post's old Torch code base. In the process, I have learned PyTorch and have trained character-level RNN language models that "hallucinated" prose of seminal Polish writers such as Henryk Sienkiewicz. Of course, pre-processing the authors' raw texts, so they could be fed as input to the NNs constituted half if not more of the work.

Later I polished my PyTorch skills during university courses. I have used PyTorch in **Computer Vision (CV)**, **Reinforcement Learning (RL)** and **Natural Language Processing (NLP)**. In NLP I have also worked with **spaCy** and the **Transformers** library of **Hugging Face**. My most ambitious deep learning project to date is titled: ***Is RGB all you need? The usefulness of depth in semantic segmentation***. Please check its devoted section for more details.

**Web Development**

JavaScript is my second language of choice after Python. I have experience in **full-stack** development. In front-end, I have worked with vanilla JavaScript enhanced by the DataTables library at Almar IT. On the other hand, at Konsept I have used React. The Python Flask library is my go-to choice for the server side when building web apps. Yet, at Almar IT, I have also used the Java Spring Boot framework in the development of a CRM based on the MVC architecture.

### Natural Language Processing

In the field of **Natural Language Processing (NLP)**, my university has offered me multiple courses such as [Processing Formal and Natural Languages](http://www.drps.ed.ac.uk/18-19/dpt/cxinfr08008.htm), [Text Technologies for Data Science](http://www.drps.ed.ac.uk/20-21/dpt/cxinfr11145.htm), [Accelerated Natural Language Processing](http://www.drps.ed.ac.uk/21-22/dpt/cxinfr11125.htm), [Natural Language Understanding, Generation, and Machine Translation](http://www.drps.ed.ac.uk/21-22/dpt/cxinfr11157.htm) (NLU+). As a result, I have extensively studied NLP basics and specialised methods devised to solve the discipline's subproblems. Especially, in the NLU+ course, we have explored the state-of-the-art approaches to NLP, such as attentional models, transformers, BERT or even the brand new pre-trained model prompting paradigm. In practical terms, I have utilised PyTorch with the modern spaCy and Hugging Face's Transformers libraries. In the end, it is worth mentioning that I had the privilege to learn from the world's top researchers in the field. For example, professor Alexandra Birch, one of the inventors of the byte-pair encoding (BPE) method used virtually everywhere in NLP for text segmentation, is also an NLU+ course lecturer.

**Databases**

MongoDB is my first database of choice. I have designed schemas, built and optimized indexes, and created queries for it. I have used MongoDB during my work at Konsept. Also as part of the [Text Technologies for Data Science](https://www.inf.ed.ac.uk/teaching/courses/tts/) course at my university, I was involved in the Lyrix project, where we created a search engine for song lyrics. The project involved information retrieval from a collection of over 2 million songs and used MongoDB in the system back-end.

I am familiar with traditional relational databases. Both as part of my university courses and my work at Almar IT, I had to design and instantiate table schemas, and write SQL queries.

**Version Control Systems**

I have used Git as VCS in my own and university group projects.

**UNIX**

The computing infrastructure at my university was based on Linux machines, so I had to learn UNIX and GNU Bash script. I use Windows on my personal machines.

**Virtualization**

When I must use a library not supported on Windows, I use Vagrant with VirtualBox to set up a VM with appropriate dependencies. In such scenarios involving Python, I like to set up a remote Jupyter Notebook server on the VM, but interact with it within the browser on my host machine. This was especially crucial for my Master's dissertation thesis project.

**Other**

The above list includes the remaining programming languages I have used in my university projects.

As a side note, Haskell was the first language I learned, and to this day I have a good grasp of the functional programming paradigm. I have also attended programming tutorials with professor Philip Wadler, one of the main creators of the Haskell language.

**Blockchain Technologies**

In my final academic year, I have undertaken a Blockchains and Distributed Ledgers course. In the course, I have studied the core principles behind the technology, and its instantiations in the form of Bitcoin and Ethereum. In practical terms, I have designed, implemented in Solidity, optimized for security, fairness and gas fees, deployed, and interacted with Ethereum smart contracts. My final A1 (95%) grade in the course shall inform one about my *"Crypto"* concepts comprehension level. Finally, inspired by the course I have implemented a distributed version of the game of chess, using the Ethereum smart contracts.

**Experience**

**Konsept App**

Konsept by [Nagne Studios](https://www.nagne.io/) start-up is a project aimed at delivering a web application that simplifies and automates the recurring parts of the product manager's (PM) job. As the project was early in development, I was directly involved in designing app features and detailing their implementation possibilities, given my machine learning background. To guide my design decisions and understand the PM's perspective I have taken the time to observe and investigate the work of Nagne Studio's PMs. A major feature of the Konsept App in the design of which I was involved concerned a recommender system. To create such a system, we had to build a database of items for recommendation. I have gathered such a dataset by implementing a custom web crawler and data parser. The next step concerned data annotation, and for this purpose, I have built a proprietary web application. Concerning my programming involvement, I have used Python, Flask, React, JavaScript and MongoDB during my work at Konsept.

**Tartan Weaving Mill**

Tartan Weaving Mill is the largest souvenir shop in Edinburgh. In the shop, I was directly involved in customer service.

**Almar IT**

[Almar IT](https://www.almar.pl/) is a medium-sized company delivering CRM systems to petroleum distributing businesses. During my internship, I was introduced to the company's large code base. The involved technological stack primarily included Java, Spring Boot, Hibernate, PostgreSQL/Oracle database, JavaScript, and DataTables library. In order to contribute, I had to learn Spring Boot and JavaScript from scratch. Finally, I was involved in developing a feature of the main CRM product.

**Languages**

**Polish**

I am a native speaker of the Polish language.

**English**

Since high school, all of my education was undertaken in English.

I have lived, studied and worked in the UK for 5 years.

My grade 6 in English B HL of the International Baccalaureate Diploma is a certificate of my English language proficiency at the B2+ level ([reference](https://www.ibo.org/news/news-about-the-ib/benchmarking-diploma-programme-language-courses-to-the-cefr/)).

**German**

I have learned German for 4 years and hold a Goethe-Zertifikat A1: Fit in Deutsch 1.

**Projects**

**Is RGB all you need? The usefulness of depth in semantic segmentation**

As part of the [Machine Learning Practical](http://www.drps.ed.ac.uk/20-21/dpt/cxinfr11132.htm) course, together with [Maciej Kowalski](https://mkowalski.me/), we have approached a **Computer Vision (CV)** problem of utilising non-RGB data to improve the performance of existing classifiers. Specifically, we have attempted to outperform the RGB-only baseline model on the dataset of drone-captured aerial landscapes, by incorporating the also available per-pixel depth/elevation information recorded with the LIDAR sensors. The task concerned semantic segmentation, so assigning a label to each pixel of an image, to determine whether such a pixel is a part of a house, road, car or tree, etc.

We have investigated different ways of augmenting the existing ResNet backbone architecture to add depth information. We have also aimed to utilise the off-the-shelf pre-trained models rather than training new ones from scratch. A meaningful introduction of untrained data channels into already pre-trained architecture posed a grand challenge. In the project, we had to schedule our hyper-parameter tuning experiments on the available university GPU cluster. Additionally, one of the solutions required me to overwrite the existing convolution operation to make it "depth-aware". In the process I have implemented a custom PyTorch convolution module, using basic vectorized mathematical operations. The module is available on my [GitHub](https://github.com/mrybok/DConv).

The final project results can be found in the attached pdf report.

**Free-time**

At the moment, when I am not working you will find me training at the gym, listening to Lex Fridman Podcast, reading Stanisław Lem's sci-fi novels or playing Elden Ring.

**Contact**

[marcinrybok@protonmail.com](mailto:marcinrybok@protonmail.com)