

## Árpád Goretity

Data Scientist, Software Solutions Architect



[arpad.goretity@gmail.com](mailto:arpad.goretity@gmail.com)



[github.com/H2C03](https://github.com/H2C03)



[linkedin.com/in/h2co3](https://linkedin.com/in/h2co3)



## Education

- **PhD in Computer Science** — 2020–2024 (*in progress*)  
Pázmány Péter Catholic University, Faculty of Information Technology and Bionics  
*Research Topic:* Design and Implementation of a Strongly-Typed Data Abstraction Layer and Domain-Specific Language
- **MSc in Data Science** — 2018–2020  
Università degli Studi di Padova, Dipartimento di Matematica “Tullio Levi-Civita”  
*Thesis:* Towards Personalized Disease Risk Prediction from Metagenome Analysis of the Microbiome  
*Qualification:* 110 / 110, Cum Laude
- **BSc in Molecular Bionic Engineering** — 2013–2016  
Pázmány Péter Catholic University, Faculty of Information Technology and Bionics  
*Thesis:* Design and FPGA Implementation of a Protein Structure Comparison Method Based on Alignment of Backbone Conformations  
*Qualification:* Excellent



## Professional Experience

- **Bioinformatics Scientist** — Medipredict Ltd., 2020–current
  - Analysis of clients' microbial and metabolomic data via classical statistical methods
  - Training and evaluation of predictive machine learning models for rare diseases, with special focus on small *in vivo* sample sizes
  - Authoring tools for streamlining work of omics scientists, for management of big data, preprocessing, QC, and cleaning of raw data files (genetic and metagenomic); containerization of research-grade software; implementation of novel algorithms
  - Development of reproducible and standardized statistical protocols to be used with in-house analytical (e.g. LC-MS) processes
  - Performing code review for other bioinformaticians
  - Mentoring interns and trainees with regards to software engineering good practice and data analysis methodology
  - Advising omics scientists on mathematical treatment of domain-specific data

Principal technologies, languages, and libraries involved:

- Python 3
- Rust
- NumPy, Pandas, statsmodels, scikit-learn, Matplotlib, Seaborn
- Docker

- **Data Platform Architect** — S+H Portfolio PLC, 2020–current

- Design and implementation of the automated analysis and diagnostic framework of the “PrivátDoktor” personalized telemedicine platform from scratch. This software suite helps GPs focus on suspicious cases and obviates the need for continuous human attendance on part of the physician on call.
- Fitting, tuning, and quality control of machine learning models based on ECG (electrocardiogram) and PPG (photoplethysmogram) signals, for the detection of common heart diseases, such as Arrhythmias, Atrial Flutter, Atrial Fibrillation, Infarction, Bundle Branch Block, and AV Block. This task required extensive application of novel feature engineering methods to time series.
- Development of an unsupervised on-line learning method for creating personalized physiological profiles for patients based on historical measurements and the detection of anomalies in newly-acquired data.
- Refactoring of inherited signal processing and analytical code for ensuring correctness and improving maintainability.
- Documentation of existing and freshly-developed machine learning and diagnosis automation methods for meeting regulatory requirements.
- Consulting with medical doctors and other healthcare specialists, mediating between upper management and the scientific advisory board, so as to reconcile business goals with high professional standards.

Principal technologies, languages, and libraries involved:

- Python 3
- Rust
- TypeScript, Node.JS
- SQLite
- NumPy, Pandas, scikit-learn, sktime, Matplotlib, Seaborn
- Vue.js 3, WebAssembly
- Docker
- REST APIs

- **Site Reliability Engineer** — NumberEight Technologies Ltd., May–July 2022

- Gradually moving deployment of existing webservice to Terraform
- Deploying Google Managed Prometheus in a fully-automated manner, following modern Infrastructure-as-Code (IaC) principles
- Creating a one-click setup for a containerized development environment to be used by future SREs
- High-level planning of operational principles and technical goals of logging and monitoring infrastructure

Principal technologies, languages, and libraries involved:

- TypeScript, Node.JS
- Docker, Kubernetes, Terraform
- GitLab CI, GitLab Managed Terraform
- Google Cloud Platform, Google Managed Prometheus

- **iOS Application Developer** — Pay with Curl Ltd., 2017–2018  
Principal technologies, languages, and libraries involved:
  - iOS, Objective-C, Swift
  - UIKit, Push Notifications
  - iBeacon
  - Xcode, Interface Builder
- **iOS Application Developer** — iCsekk/Díjbeszed IT PLC, 2016–2017
- **iOS Application Developer** — Codie (ex-RoboPhone), April–August 2014
- **iOS Application Developer** — OTP SmartBank/EQL Soft Ltd., June–August 2012



## Other Technical Skills

- Proficiency in systems programming using modern C and C++
- Proficiency with database management, data modelling, schema design, and query optimization in SQL (mainly SQLite and Postgres) and MongoDB
- Familiarity with embedded platforms, proficiency with Atmel AVR
- Some exposure to Haskell, intermediate-level knowledge of pure functional programming
- Some exposure to FPGA design, VHDL programming, and high-level synthesis (HLS) from C and C++ sources
- Unix shell scripting under Linux and macOS
- Source Code Management using Git, familiarity with GitHub and GitLab
- Technical typesetting with L<sup>A</sup>T<sub>E</sub>X
- Teaching experience (TA, later Grad Student Instructor @ PPCU FITB):
  - Linear Algebra I-II — 2014/15/1–2
  - Nonlinear Dynamical Systems in Biology — 2020/21/1
  - Introduction to Programming I — 2015/16/1, 2017/18/1
  - Introduction to Programming II — 2020/21/2
  - Programming I — 2014/15/2
  - Parallel Programming — 2020/21/2



## Publications

- Árpád, GORETITY; Zoltán, NAGY; Zoltán, GÁSPÁRI;  
“Acceleration of a Protein Structure Comparison Algorithm on FPGA,”  
in *2017 European Conference on Circuit Theory and Design (ECCTD)*, 2017 [online].  
DOI: 10.1109/ECCTD.2017.8093342
- Árpád, GORETITY; István, REGULY;  
“Query Complexity in Modern Database DSLs,”  
*manuscript, under peer review*



## Talks and Presentations

- **Compiler Design and Implementation**

*Swift Meetup Budapest, September 2015–June 2016*

A 6-part, practice-oriented series, in which participants implement a simple, working compiler in Swift for a procedural toy programming language.

- **Non-Pessimizations in the Swift Compiler**

*Swift London, 15 November 2016*

A discussion of interesting canonicalization and optimization patterns, as implemented by the Swift compiler.

- **Acceleration of a Protein Structural Alignment Method with FPGAs**

*Budapest Science Meetup, 20 February 2017*

A simplified recount of my Bachelor's thesis, discussing the significance of hardware acceleration in bioinformatics and big data.



## Language Proficiency

- **English** — Cambridge Certificate of Proficiency (Level C2)
- **French** — Level C1
- **Italian** — Basic familiarity
- **Hungarian** — Native speaker