

# TOMÁŠ FRÝDA

+420 723 089 475  
me@freedatoms.eu  
freedatoms.eu

Svatovítská 361/4, Prague 6, Czechia, 160 00

**GitHub:** freedatoms, tomasfryda

**LinkedIn:** freedatoms

## WORK

**2020 – now** I joined H2O.ai to work on open-source AutoML and Stacked Ensembles in H2O-3, focusing on Java, Python, and R.

Thanks to my former manager, Erin Ledell, PhD, I developed a deeper interest in explainability, interpretability, and fairness in machine learning. This led me to create the explain module and to implement SHAP for most supervised ML algorithms in H2O-3 (described here).

I subsequently worked on interpretable models, including AdmissibleML and GLM, where I implemented a Tweedie family estimator with simultaneous estimation of dispersion and Tweedie variance power — a particularly enjoyable challenge (PR 1, PR 2). I also became the H2O-3 R package maintainer.

When H2O.ai pivoted toward Generative AI, I joined the team behind H2O Eval Studio due to my continued interest in the explainability and interpretability of ML models. I was responsible for the machine learning components of the product, including the evaluation of large language models (LLMs), retrieval-augmented generation (RAG), and agentic systems.

Some of my work on H2O Eval Studio was open-sourced in the H2O Sonar package, namely the GenAI Evaluators.

My official job title is *Senior Software Engineer*.

**2018 – 2020** As a freelance *Data Scientist*, I primarily worked in R and Python across various projects.

I spent most of my time collaborating with Datamole, where I started as a *Data Scientist* and was soon promoted to *Data Science Team Lead*. My responsibilities included developing optimization procedures, contributing to creation of internal MLOps system, and assisting in the design and implementation of a data warehouse.

Beyond Datamole, I was involved in teaching Distributed Data Mining at the Czech Technical University and provided consulting services in data science, information theory, statistics, and graph theory.

**2017 – 2018** I joined Vendavo as a *Data Scientist*, focusing on price optimization in R. This role offered a highly enriching environment where I had the opportunity to collaborate with a diverse and experienced team. I worked alongside Petra Dědičová, PhD (Econometrics), Miroslav Čepěk, PhD (Artificial Intelligence), and my manager Luděk Kopáček, PhD, who brought a rare combination of deep expertise in both economics and AI. Their influence pushed me to deepen my understanding of classical statistical methods and broaden my perspective on pricing strategies and economic modeling.

**Jul 2016 – Sep 2016** I took part in the Research Summer at FIT (VýLeT) at Czech Technical University, where I selected my research topic: scalable machine learning of predictive ensembles. My work focused on using the h2o-3 platform and the FAKE GAME approach (Fully Automated Knowledge Extraction using Group of Adaptive Models Evolution). I'm especially grateful to my supervisor, Pavel Kordík, PhD, for his guidance and support throughout the project. This research later evolved into my master's thesis and contributed to several publications.

## EDUCATION

**2014 – 2017** Knowledge Engineering, Faculty of Information Technology, Czech Technical University

**Master's thesis:** Scalability of Predictive Modeling Algorithms which received Dean's Award in Winter semester 2016/2017

**2011 – 2014** Theoretical Computer Science, Faculty of Information Technology, Czech Technical University

**Bachelor's thesis:** Comparison of different neuroevolution algorithms on several problem domains

**2010 – 2011** Mathematical Informatics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University

Unfinished; transferred to Faculty of Information Technology for better alignment with interests

## INTERESTS & HOBBIES

---

### Linux & Open Source

I first began using Linux around 2001 with Debian Potato, and over the years I explored a variety of distributions – Mandrake, Mandriva, OpenSUSE, Ubuntu, and Arch. When Arch announced its transition to SystemD, I moved to Gentoo, which has remained my primary environment at home (with occasional experiments in more niche distributions like GNU/Guix). Along the way, I also explored FreeBSD and OpenSolaris. While at university, I used Ubuntu for stability in my coursework and Gentoo for deeper hands-on experimentation.

My open source work includes forking the Arora browser to create Voyager, contributing Czech translations for the Editra text editor, and more recently working on the H2O-3 machine learning package as part of my professional work. I also had a small footprint in the Arch Linux community by maintaining several PKGBUILDs in the AUR – customized GNU Emacs, MKCL (a Common Lisp implementation with strong threading support at the time), SBCL (to replace the unstable official build prone to segmentation faults), and rekonq-git.

Since its inception in 2012, I have attended the Linux Days conference almost every year, except during the COVID-19 years.

### Statistics & Machine Learning & AI

I've always been a curious person, so it felt natural to ask myself how to best satisfy that curiosity. For me, the answer has been the scientific method, with statistics as one of its essential tools. My fascination with computers began in early childhood (when, instead of going to pre-school, I was often watching Star Trek with Lt. Cmdr. Data). The combination of mathematics, statistics, and computing eventually led me to Machine Learning. My curiosity isn't limited to ML alone — I also enjoy thinking about broader areas of AI, such as game theory, and I have a particular interest in probabilistic programming languages. I started with Figaro (inspired by Practical Probabilistic Programming from Manning) and later explored Anglican, Stan, and PyMC. More recently, I've been looking into conformal prediction for uncertainty estimation, which I find more generally applicable to machine learning than traditional Bayesian methods.

Within ML, my interests include Automated Machine Learning and Neuroevolution, both of which formed substantial parts of my academic theses. Staying true to my inquisitive nature, I also gravitate toward interpretable models and methods that help explain and demystify the behavior of machine learning systems.

Being dyslexic has shaped my perspective, making me more empathetic toward marginalized groups and more engaged in topics of fairness, accountability, transparency, and the prevention of algorithmic discrimination.

### Cooking & Nutrition

I'm a tolerant person—except when it comes to lactose, which I can only handle in small doses. Finding out I was lactose intolerant made me realize just how omnipresent lactose is, which led me to start cooking for myself. Being a geek, I began with Cooking for Geeks, which opened a culinary rabbit hole that eventually led me to Hervé This's Kitchen Mysteries and other works by This and Nicholas Kurti, the pioneers of what we now call molecular gastronomy. From there, I ventured into Modernist Cuisine at Home, and later my curiosity shifted toward the flavors, techniques, and traditions of Mediterranean cuisine. By now, I probably own more books on culinary science than on machine learning.

My interest in food isn't just about technique — it's also about health. Wanting to improve my diet and lose weight, I followed a recommendation from Hacker News to read Michael Greger's How Not to Die, a book I particularly appreciated because it cites a scientific paper for every single claim it makes (though it leans toward a vegan perspective). This inspired me to attend ProVeg's plant-based nutritionist courses for two consecutive years. In theory, I now know exactly what I should eat, but in practice my eating habits still occasionally diverge from that ideal.

### My dogs Pixel & Okurka

Outside of work, I spend most of my time with my dogs. My older dog Pixel is a bright and energetic Australian Shepherd who quickly adapted to me working from home during the COVID years and Okurka — a Shih-tzu puppy doesn't know any other way. As a result, remote work or a nearby dog-friendly office has become a strong plus for me.

## PUBLICATIONS

---

I try to keep up-to-date ORCID, I also have a Google Scholar profile.

- Nykodym, Tomas and Kraljevic, Tom and Wang, Amy and Wong, Wendy and Fryda, Tomas (2025). *Generalized linear modeling with h2o*, Published by H2O.ai Inc. Available here.
- Kordík, Pavel, Černý, Jan, Frýda, Tomáš. (2018). *Discovering predictive ensembles for transfer learning and meta-learning*. Machine Learning. 107. doi: 10.1007/s10994-017-5682-0.
- Kordík, Pavel, Frýda, Tomáš (2018). *On Scalability of Predictive Ensembles and Tradeoff Between Their Training Time and Accuracy*. In: Shakhovska, N., Stepashko, V. (eds) *Advances in Intelligent Systems and Computing II*. CSIT 2017. *Advances in Intelligent Systems and Computing*, vol 689. Springer, Cham. doi: 10.1007/978-3-319-70581-1\_18.
- Kordík, Pavel, Frýda, Tomáš, Šnorek, Miroslav, Čepek, Miroslav. *Scalability of predictive ensembles*, 2017 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), Lviv, Ukraine, 2017, pp. 555-560, doi: 10.1109/STC-CSIT.2017.8098848.
- Frýda, Tomáš (2017). *Scalability of Predictive Modeling Algorithms*, České vysoké učení technické v Praze. Available here.
- Frýda, Tomáš (2014). *Comparison of Different Neuroevolution Algorithms on Several Problem Domains*, České vysoké učení technické v Praze. Available here.