



## Seminar Hora Informaticae

Institute of Computer Science, Prague

Tuesday, May 20, 2025, 13.30 – 15.30 (1.30 – 3:30 PM) CEST

Meeting Room 318, Address: Pod Vodárenskou věží 2, Prague 8

Meeting ID: 914 0834 4018, Passcode: 668534



<https://cesnet.zoom.us/j/91408344018?pwd=x2Qlz4F42BxlMSmWc1HOwHHA7Uw7PN.1>

---

**Tomáš Pevný, Czech Technical University:**

### **Understanding the neural networks through rule extraction.**

Neural networks are ubiquitous yet they remain opaque for most of its users, who has very little understanding of how they store the knowledge and how the information propagates through. In this talk, I would like to share our findings from our quest to understand these phenomena. Specifically I will show the decision rules realized by neural networks and why it might be difficult to understand them without the knowledge of the data distribution. This will give us intuition why neural networks are robust yet why adversarial samples are so easy to create. Finally, we will use these tools to understand, how the decision rules compose during inference.

---

**Tomáš Pevný** (<https://cs.felk.cvut.cz/en/people/pevnytom>), born in 1979, is a researcher and lecturer at Czech Technical University in Prague. He received his PhD in Computer Sciences from State University of New York in Binghamton in Computer Science in 2008 and an MS in Computer Sciences from Czech Technical University in Prague in 2003. In 2008–2009, T. Pevny spent one year in Gipsa-lab in Grenoble, France, as a post-doctoral researcher. From 2013-2019, he was partially working in Cisco as a consulting scientist. From 2019-2025, he also partially works for Gen digital Inc. (former Avast). His main research interests is in tailoring machine learning algorithms to enable their application in a security industry.

---

**HORA INFORMATICAЕ** (meaning: TIME FOR INFORMATICS) is a broad-spectrum scientific seminar devoted to all core areas of computer science and its interdisciplinary interfaces with other sciences and applied domains. Original contributions addressing classical and emerging topics are welcome. Founded by Jiří Wiedermann, the seminar is running since 1994 at the Institute of Computer Science of the Czech Academy of Sciences in Prague.

<https://www.cs.cas.cz/horainf>